FOREIGN EXCHANGE

Practical Asset Pricing and Macroeconomic Theory

ADAM S. IQBAL

Foreign Exchange

Adam S. Iqbal

Foreign Exchange

Practical Asset Pricing and Macroeconomic Theory

palgrave
macmillan

Adam S. Iqbal
London, UK

ISBN 978-3-030-93554-2 ISBN 978-3-030-93555-9 (eBook)
https://doi.org/10.1007/978-3-030-93555-9

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Mum, Dad, my wife, Gosia, my daughter, Maria, and to my sister, Sadia,
who showed me what could be done.

Preface

One of the great challenges that I and, I believe, many participants in foreign exchange (FX) markets face is sifting through the often overwhelming amount of information that is available. Media outlets stream updates on international politics, economics and other factors that move FX prices twenty-four hours a day. It is difficult to work out what is and what is not important.

In my own attempt to filter the information barrage, I cobbled together a disorganised set of ideas on the economics relating to exchange rate determination from a disparate set of sources: text books, academic literature, industry research notes, conversations with other market practitioners and theories cited in media reports. I started writing Foreign Exchange: Practical Asset Pricing and Macroeconomic Theory in the winter of 2020 not as a book, but as a set of self contained notes to myself to organise the many ideas that I have come across, to improve my ability to sift through the daily plethora of FX information with better rigour, and to be able to translate the information into concrete FX views that are firmly rooted in the economic theories of risk premiums, interest rates, and inflation, among other topics. This approach promotes time consistent thought that avoids the daily temptation to jump from that day's economic narrative to the next.

As my notebook grew, I found myself wanting to cite it in conversation and in my writing. This led me to publish. I hope that my target audience of buy- and sell-side industry practitioners, finance and economics undergraduate and graduate students, M.B.A. students, academics and others interested in FX markets will find this book helpful in their thinking.

The book proceeds in a manner similar to Iqbal (2018) in that it is inspired by academic economic theory, but its emphasis is on building economic intuition. The approach taken may prove particularly valuable to market practitioners tasked with making decisions in an environment where time constraints mean that simplicity and intuition are of greater value than mathematical formalism.

I presume some exposure to basic probability and calculus. However, this should not deter readers with less exposure to these topics for at least two reasons. First, the presentation is consciously informal with the aim of exposing ideas in their simplest form before going back and understanding their foundations. Second, I provide several mathematical appendices to assist such readers and to keep this book self-contained with respect to the most important concepts. Although at times it may not seem to be the case, the challenges in FX economics are conceptual rather than formulaic.

London, UK Adam S. Iqbal

Acknowledgements

This book owes a debt of gratitude to my present and former colleagues at Goldman Sachs, PIMCO, and Barclays Investment Bank for their engagement in our discussions on foreign exchange and foreign exchange options and for all that they have taught me. In particular, I thank Nick Wilson and Thomas Burrell for their relentless drive to make all things theoretical, practical, which has partly motivated this piece.

I also thank those who have commented on, helped review and provide suggestions, namely Johannes Ruf, Jeffrey Chen and Tony Song, among others.

A very special thank you goes to Mobeen Iqbal for his effort in proof-reading, and providing invaluable feedback.

All remaining errors are my own.

Contents

About the Author

Adam S. Iqbal is a Managing Director and Global Head of G10 FX Options Trading at Goldman Sachs,[1] where he has previously served as Global Head of FX Exotics and Correlation Options Trading, and EMEA Head of G10 FX Options Trading, running FX vanilla and exotic options portfolios. Prior to this, he was an FX Volatility Portfolio Manager at PIMCO, and he has worked as a vanilla and exotic FX options trader at Barclays Investment Bank in London. He is also author of *Volatility: Practical Options Theory* (2018).

Dr. Iqbal holds a Ph.D. in financial mathematics and economics from Imperial College London, an M.Sc. in applied mathematics from Oxford University, and an M.Sci., and B.A. in physics from Cambridge University. He has held a visiting academic position at Imperial College London and has guest lectured at the London School of Economics.

[1] This work represents the author in a personal capacity only, and does not contain any views or opinions of Goldman Sachs.

Part I

Risk Premiums and the Central Pricing Equation

Presented with an opportunity to trade an actuarially fairly priced security, by which I mean that its price today is equal to its objective expected payoff, a rational investor should choose based on which side of the trade best hedges her portfolio. The reason is that this side decreases her portfolio variance, without decreasing her expected return. However, since investor portfolios are correlated to each other in aggregate, more will bet the same way as each other than opposing each other, leading to unequal capital flows at the actuarial price. This flow imbalance then drives the market price of the security away from its actuarial value. The difference that opens up between the market price of the security and its actuarial value is known as the risk premium. The risk premium is then earned over time as compensation to investors taking the side of the trade that adds to aggregate risk and it is paid for by those taking the side that hedges aggregate risk.

This brief description may pique the reader's interest in the concept of a risk premium and its impact on asset pricing. However, it should be noted that the risk premium is far more than just an interesting concept. Changes in the risk premium are the primary driver of price movements in FX and other asset markets[1] and are therefore more important in understanding market price movements than changes in macroeconomic variables, such as the expected real exchange rate, expected inflation, the balance of payments, interest rate differentials and the remaining plethora of factors that are cited in FX analysis. I return to macroeconomic variables and models in Part II of this book.

[1] See, for example, Cochrane (2011) for a full survey.

Almost all practitioners are at least implicitly familiar with this idea, often informally referring to falling risk premiums as *risk-on* and rising risk premiums as *risk-off* sentiment, when explaining price movements. They develop intuition for how a change in the risk premium impacts various currencies and their behaviour. The purpose of Part I of this book is to explain the concept of the risk premium both intuitively and formally, so as to facilitate a more precise understanding of its impact in FX investing.

Chapter 1 provides three carefully selected examples to build intuition around the risk premium. It goes on to formalise the idea using the utility model of investor behaviour, and derives the *central pricing equation* that I then use for FX determination over the course of the remainder of this part of the book. I also show how risk premiums enter into derivative pricing and more generally into continuous time financial models.

Chapters 2 and 3 incorporate the risk premium into common approaches to FX investment. Chapter 2 focuses on covered interest parity (CIP), uncovered interest carry (UIP) and the carry trade. I show that the expected profits from carry trading result from the risk premium, and that it is a distinct concept to *carry* itself, which should rather be thought of as a buffer against adverse price movements.

Chapter 3 describes a common approach to FX determination. It shows how to combine the law of one price/purchasing power parity model of FX, inflation expectations and CIP to determine the FX rate. It also shows how to *add in* the risk premium into this otherwise relative standard form of FX analysis.

Finally, Chapter 4 discusses FX volatility. I introduce FX options as instruments to trade volatility risk. I discuss how their very existence implies that volatility is itself uncertain, and that this gives rise to another risk premium, the so-called volatility risk premium. I also apply the central pricing equation of Chapter 1 to derive the intuitive result that volatility in the real FX rate arises from imperfect international risk sharing. If risks were equally shared across foreign and domestic investors, then there is no need for the real FX rate to ever move.

1

Risk Premiums

Through daily experience, market practitioners are familiar with government bond, currency and other asset prices exhibiting strong correlations with investor sentiment and with the global economy.

In the government bond markets of developed economies, negative economic news typically causes yields to fall. It is perhaps counter intuitive that the price of government debt rises during recessions, just as its fiscal position weakens due to falling tax receipts and rising unemployment benefit claims. Nevertheless, the correlation between developed market government bonds and the state of the global economy remains negative.

In the currency markets, emerging market currencies, and some developed currencies, such as the Australian Dollar (AUD) and Norwegian Krone (NOK), tend to fall in value relative to so-called safe-haven currencies such as the U.S. Dollar (USD), Swiss Franc (CHF) and Japanese Yen (JPY). Perhaps counter-intuitively, such price movements occur even if the negative economic news emanates from the United States (U.S.) itself! Consider, for example, the strength observed in the USD in the aftermath of the meltdown of the U.S. housing market and collapse of Lehman Brothers, among many other economic difficulties in 2008. Similar strength in the USD was observed in 2011, in the aftermath of the debt ceiling crisis, when Standard and Poor's (S&P) downgraded U.S. government debt from AAA to AA+.

Such market behaviour suggests that, like government bonds, a substantial component of daily price movements in currencies is not driven by the economic fundamentals of the issuing country. It is, quite simply, incorrect

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_1

to view negative (positive) news from the U.S. as negative (positive) for the USD, just as it is incorrect to view negative news about the U.S. economy as negative for U.S. government bonds.

The fact that government bonds and certain currencies exhibit negative correlations with the global economy and others exhibit positive correlations must effect their pricing today. One may intuit that investors may pay a premium to hold government bonds and safe-haven currencies, and demand a discount to hold riskier currencies. This is the so-called *risk premium.*

This chapter formalises our understanding of the risk premium, and in doing this, the reader will be able to better understand the economic drivers of day-to-day price movements in currencies, government bond yields and other assets.

1.1 Risk Premiums

The actuarial value of an asset at time t, V_t^P, is given by,

$$V_t^P = \mathbb{E}_t^P[X_{t+1}], \tag{1.1}$$

where X_{t+1} denotes its random payoff of in the next period, and $\mathbb{E}_t^P$ denotes the time t conditional expectation (see Appendix B.1 for a refresher on calculating expectations). I assume interest rates are zero until later in the chapter and there is therefore no time discounting at this stage.

The probabilities that the expectation in Eq. (1.1) are calculated under reflect the actual, or objective probabilities associated with the payoff. Henceforth, I refer to these as P probabilities. For example, a fair coin toss that pays \$1 if a coin lands on heads, and nothing if the coin lands on tails is priced using the P probabilities of P(Heads) $= 0.5$ and P(Tails) $= 0.5$ to give a $V_t^P = \$0.5$.

The *risk premium* is a concept in financial economics that expresses the idea that assets are not priced at their actuarial value. The market price of the asset, V_t, is given by

$$V_t = \mathbb{E}_t^Q[X_{t+1}], \tag{1.2}$$

where $\mathbb{E}_t^Q$ is the expected payoff calculated under a so-called *risk-neutral* or *risk-adjusted* probability measure. Henceforth, I refer to these probabilities as Q probabilities.

Pricing under the Q probabilities leads to assets trading at either a premium or a discount relative to their actuarial value. In the coin toss example, the price to partake in the bet, V_t, may be more or less than \$0.5 depending on the Q probabilities that the market assigns to each of the outcomes. The purpose of the rest of this chapter is to make clear the meaning of $\mathbb{E}_t^Q$ and the risk premium.

Before exploring the microeconomic foundations and mathematics underlying risk premiums, I build intuition through three selected examples, (i) pricing a risky corporate bond, (ii) a bookmaker setting odds on a national election, and (iii) pricing a one-period call option on GBP-USD. Each example places emphasis on a different area of this important topic.

In example (i) I ask the reader to take the existence of the risk premium as given so that I can demonstrate its meaning in arguably its simplest form, using a one-period, two-state model. I make some informal suggestions as to why risk premiums may exist.

Example (ii) shows how risk premiums naturally arise via capital flows. I show that a bookmaker must set his odds to respect the flows of capital betting on the election outcome and that, perhaps counter-intuitively, he can run his business without paying much attention to the objective, P, probabilities of the event upon which he is accepting bets. As a corollary to this example, I discuss how changes in risk premiums can lead to market volatility that may appear to be irrational to an observer of the objective probabilities, but are actually the result of rational portfolio decisions by investors.

Finally, example (iii) discusses risk premiums in the context of derivatives. I show that risk premiums are consistent across assets and their derivatives. This example also provides an introduction to the relationship between dynamic hedging arguments for pricing derivatives first introduced by Black and Scholes (1973) and Merton (1974) and the financial economic theory of risk premiums.

1.1.1 Example (i): A Risky Corporate Bond

Consider a risky corporate bond that will pay the investor \$1 if the company survives, and nothing if it fails. Suppose also that market participants have agreed that the company's objective default probability is 10%. The actuarial value of the bond is calculated as

$$
\begin{aligned}
V_t^P &\equiv E_t^P[X_{t+1}] \\
&= \mathrm{P(Survives)} \times \$1 + \mathrm{P(Fails)} \times 0 \\
&= 0.9 \times \$1 + 0.1 \times 0 \\
&= \$0.9.
\end{aligned}
$$

If there is a risk premium associated with this bond, the market may allow the investor to purchase it for a discounted price of, say, \$0.8, relative to its actuarially fair value of \$0.9. Instead of pricing using the objective P probabilities, the market prices the bond using Q probabilities as follows.

$$\begin{aligned} V_t &= E_t^Q[X_{t+1}] \\ &= \text{Q(Survives)} \times \$1 + \text{Q(Fails)} \times 0 \\ &= \$0.8. \end{aligned} \tag{1.3}$$

In words, even though market participants know that the company has a 90% probability of surviving and paying out \$1, they trade it at a price that implies a 80% probability of survival. The buyer of the bond has an expected payoff of \$0.1 (log-return of 11.7%). This is her *risk premium.* The idea of pricing the asset using the Q probability distribution rather than the P probability distribution is known in finance literature as a *change of measure*, *risk-neutral* pricing or *risk-adjusted* pricing among several other synonymous terms. Note that in this example, the P probabilities are obtained from objective analysis of the company and economy, but the Q probabilities are inferred from its market price in that Q(Survives) = 0.8 was inferred from Eq. (1.3) and the traded price of \$0.8.

One may naturally ask, why would the market allow the \$0.1 discount? Would investors seeking positive expected returns not buy up the bond to push its value back up to \$0.9? We will see more formally later in the chapter that modern financial economic theory tells us that the answer is generally no, because the risks associated with financial assets are not fully diversifiable. At this stage, I provide some intuition.

Suppose that the bond is highly correlated with the wider economy. The economy has a 10% probability of entering a recession, and in this circumstance the corporate defaults. If the economy does not enter a recession, then the bond survives. This is consistent with an actuarial value of the bond of $V_t^P = \$0.9$.

Since the bond will default in a recession, when the remaining assets in the investor's portfolio are also likely to perform poorly, and it will survive in a healthy economy when the remaining assets in the investor's portfolio perform well, the bond adds risk to the investor's portfolio rather than diversifying it. This leads the investor to demand a *risk premium* to hold the bond. The Q probabilities used by the investor to price the bond are therefore more conservative with respect to its performance than the P probabilities.

Consistent with this idea, assets that act as a portfolio hedge carry a negative risk premium. The credit default swap (CDS) on the above corporate bond is a natural example. The portfolio of the corporate bond and the CDS pay \$1 with certainty and must therefore be priced at \$1 today to prevent arbitrage opportunities. The market price of the CDS is therefore \$1 minus the price of the corporate bond, which is \$0.2 in our example. Since the CDS's expected payoff is $0.1 \times \$1 + 0.9 \times \$0 = \$0.1$, it is clear that this instrument carries a negative risk premium of $-\$0.1$.

There are several other assets that act as a portfolio hedge. Developed market government bonds have been good examples in recent times, as central banks have typically cut interest rates during economic downturns, thereby raising the prices of bonds during recessions.[1] The feature box provides examples from the FX market of currencies that may exhibit either positive or negative risk premiums based on the intuition that we have developed here.

Risk Premiums in Currencies

The U.S. Dollar (USD), Swiss Franc (CHF), Japanese Yen (JPY) and, at times, the Euro (EUR) have historically acted as portfolio hedges. However, currencies such as the Australian (AUD), Mexican Peso (MXN) and other often smaller developed market and emerging market currencies exhibit the opposite behaviour. Figure 1.1 illustrates this in the context of the negative global economic shock associated with the Coronavirus pandemic in 2020.

Based on the intuition that we have built, we should expect the USD, CHF, JPY and EUR to carry a smaller or negative risk premium because they perform well during economic turmoil and act as portfolio hedges. Similarly, AUD and MXN should carry a positive risk premium because they perform poorly during economic turmoil and add risk to a investor's portfolio.

An important question is *why* have USD, CHF, JPY and, at times, EUR, acted as portfolio hedges, while most other currencies add risk to portfolios? I discuss this topic in depth in Sect. 2.3.1.

[1] At the time of writing, many developed market central banks are nearing their effective lower bounds in interest rate policy, and so the degree to which government bonds continue to act as portfolio hedges remains to be seen.

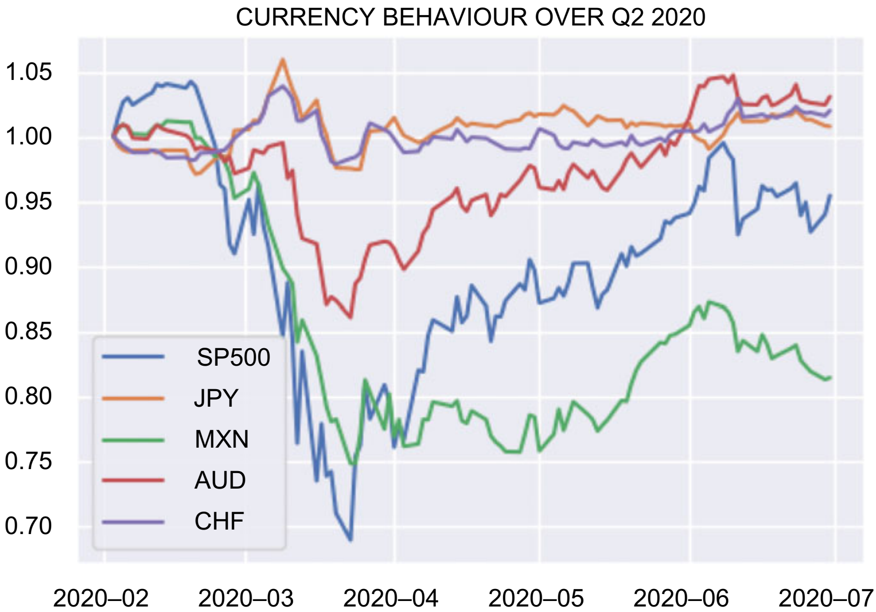


Fig. 1.1 The figure shows the behaviour of the JPY, MXN, AUD and CHF over Q2 of 2020 as the Coronavirus pandemic shook economies around the world. I also show the S&P500 stock market index. The units of the data are in US$ per unit of the currency, normalised to 1.0 at the start of February 2020. We see that as equity prices fell, so did AUD and MXN against the USD. However, CHF and JPY, initially at least, strengthened against the USD (*Source* Board of Governors of the Federal Reserve System (US), Equity Prices and Foreign Exchange Rates [SP500, DEXJPUS, DEXMXUS, DEXUSAL, DEXSZUS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/, October 1, 2020)

1.1.2 Example (ii): Bookmaker Odds

Consider the position of a bookmaker taking bets from punters on whether a national election will be won by the Economic Growth Party (EGP), or the Economic Decimation Party (EDP). Instead of pre-announcing odds, the bookmaker receives wagers, with a view to announcing odds ex post such that he is perfectly hedged.

The bookmaker receives \$70 of bets on EDP, and \$30 on EGP. With this knowledge, he assigns probabilities of 0.7 that the EDP will win, and 0.3 that the EGP will win. These probabilities correspond to decimal odds of 1.42 for EDP and 3.33 for EGP, meaning that a \$1 bet placed on EDP returns the punter \$1.42, a profit of \$0.42 and a \$1 bet on EGP returns the punter \$3.33, a profit of \$2.33.[2]

Note that the bookmaker has correctly set the odds so that he is perfectly hedged. He received \$100 from punters, and will pay out $\$70 \times 1.42 = \$100$ if EDP wins, or $\$30 \times 3.33 = \$100$ if EGP wins.

At this stage the reader may think that this example is somewhat contrived. However, its importance lies in what has not yet been mentioned. We have not mentioned the objective probability of EGP or EDP winning the election. It turns out that the bookmaker did not need to know this probability in order to hedge! He just needs to know the flows of capital.

As in Sect. 1.1.1, the risk-adjusted Q probabilities are obtained from the market price. In this case Q(EDP win) = 0.7 and Q(EGP win) = 0.3 are obtained by observing (and inverting) the decimal odds provided by the bookmaker. The objective P probabilities could be anything at all. Suppose a survey of the punters revealed that P(EDP win) = P(EGP win) = 0.5. One may then ask why the P probabilities can diverge so far from Q? The answer from the standpoint of modern financial economic theory is that investors know that EDP will, as their manifesto says, decimate the economy. Investors hedge this risk by betting on EDP to win the election, and these capital flows drive the Q probabilities away from P. This creates a risk premium in the betting market that is earned by the punters taking the other side of this risky bet, namely those betting on EGP to win the election.

Extending this example provides insight into how markets can exhibit high levels of volatility ahead of political and other events.

Volatility Associated with Risk Premiums

Consider a case where investors become more fearful of EDP's economic program. Then, during a second round of bookmaking, the bookmaker may receive \$80 of bets on EDP and \$20 of bets on EGP. As before, a hedged bookmaker will assign Q(EDP win) = 0.8 (decimal odds of 1.25) and Q(EGP win) = 0.2 (decimal odds of 5). Therefore, the price of the bet may move even though the objective probability of the election's outcome remains unchanged at P(EDP win) = P(EGP win) = 0.5.

The volatility exhibited in financial asset prices may be even larger. For instance, if investors' fearfulness was driven by the EDP becoming increasingly aggressive on economic policy over the campaign, then they may further revise down their estimate for where the stock market reprices to in the event

---

[2] Readers unfamiliar with decimal odds may understand them as follows. If the probability of an event is p then the corresponding decimal odds are $1/p$ because these odds set the expected payoff of the bet, $p \times (1/p) = 1$, equal to its cost, 1, leaving an expected profit of 0. Put more succinctly, setting decimal odds to $1/p$ prices the bet at fair value.

of an EDP win and this will lower the stock market price today. However, the increasing Q(EDP win) discussed in the previous paragraph will exacerbate the stock price fall because investors simultaneously assign a higher risk-adjusted, *Q*, probability weight to this lower stock market price. One may therefore observe substantial volatility in asset prices, all while the objective probability P(EDP win) remains unchanged at 0.5.

The important point to note is that the volatility observed in the market in this example results from rational investors making portfolio decisions. The fact that betting market and asset market prices move despite the objective probabilities remaining unchanged does not necessarily require an observer to resort to *animal spirits* or *irrationality*-based explanations of the phenomenon and investors should consider risk premium-based explanations.

In practitioner parlance, a period of increased risk aversion is typically labelled *risk-off*, and a period of decreased risk aversion is called *risk-on*. Here, I marry financial economic theory to the practitioner experience. In short, so-called risk-off refers to periods during which the divergence between *Q* probabilities and *P* probabilities increases, and risk-on to periods during which this divergence decreases.[3]

1.1.3 Example (iii): Bimodal Option Pricing in a One-Period Model

The example in this section assumes familiarity with the basics of option payoffs. Readers with less familiarity with this product may skip ahead to Sect. 4.1 for an introduction, or consult Iqbal (2018) for a full treatment.

Consider the following circumstance. GBP-USD spot is trading at 1.30 one day before the Bank of England (BOE) decides whether to hike interest rates or stay on hold (see Appendix A.1 for notes on FX quotation conventions). I take as given that a hike will send GBP-USD to 1.31, and a hold will send it to 1.29. Investors' objective assessment is that P(Hike) = 0.8 and P(Hold) = 0.2.

An FX options market maker is tasked with pricing a call option with a notional of 100 million GBP and a strike price of 1.30. She inspects the payoffs of the option (see Fig. 1.2), performs some time-consuming scenario analysis and eventually discovers a hedging strategy that allows her to price the option at 500 thousand USD, and leave her with no further exposure.

[3] Friedman (1953) wrote that a "billiard player made his shots as if he knew the complicated mathematical formulas that would give the optimum directions of travel." This description of positive economics is perhaps appropriate here because almost all market practitioners are familiar with changing risk premiums, often without a formal adoption of the theory presented here and elsewhere.

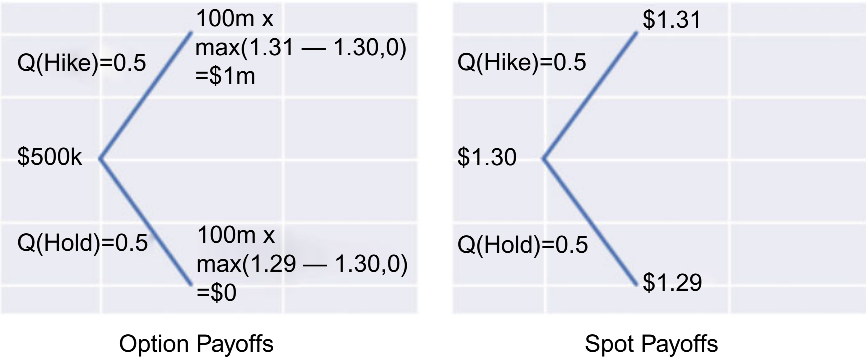


Fig. 1.2 The left plot shows payoffs of a 1.30 strike call option on GBP-USD with a notional of 100 million GBP. The right plot shows the payoffs of GBP-USD spot. Arbitrage-based pricing implies that the price of the option is 500 thousand USD. This price exactly corresponds to expected value-based pricing, but using the *Q* probabilities that implied from the price of GBP-USD spot

She proceeds with the strategy of selling the option to the investor for 500 thousand USD, and then buys GBP-USD in the amount of 50 million GBP in the market as her hedge, against selling borrowed USD.

The next day, if the BOE hike then she owes the payoff of the call option of 100 million $\times$ $(1.31-1.30) = 1$ million USD to the investor who purchased the option. However, she received 500 thousand USD in option premium, and made a profit of 50 million $\times$ (\$1.31 $-$ 1.30) $=$ 500 thousand USD on the 50 million GBP that she bought. Her profit and loss account therefore nets to zero.

If the BOE hold and GBP-USD falls to 1.29 then she owes nothing to the investor because the option payoff is zero. She received 500 thousand USD from the investor, but lost exactly this amount on the 50 million GBP that she bought because 50 million $\times$ $(1.29-1.30) = -500$ thousand USD. Again, her profit and loss account shows zero. Her hedging strategy is perfect.

Note that any deviation from a price of 500 thousand USD for the option would yield an arbitrage profit. For example, if the trader were able to sell the option for 501 thousand USD and execute the same hedging strategy then she would obtain 1 thousand USD in profit with certainty. Conversely, if the price of the option were 499 thousand USD then the buyer of the option may execute the strategy discussed above to obtain a risk-free profit of 1 thousand USD. Assuming that market pricing does not allow arbitrage, the market price of the option must be 500 thousand USD.

The interesting and important point to note is that the price that the market maker came up with of 500 thousand USD for the call option is the same as if she had simply inserted the probabilities Q(Hike) = 0.5 and Q(Hold) = 0.5 into Eq. (1.2),

$$
\begin{aligned}
V_t &= 100\text{m} \times \begin{bmatrix} \text{Q(Hike)} \times \max(1.31 - 1.30, 0) + \text{Q(Hold)} \\ \times \max(1.29 - 1.30, 0) \end{bmatrix} \\
&= 100\text{m} \times [0.5 \times 0.01 + 0.5 \times 0] \\
&= 500 \text{ thousand USD.}
\end{aligned}
$$

Could she have known of this comparably straightforward option pricing method ahead of time? If so then instead of going through the process to first identify a perfect hedge strategy and then use it to infer an option price, she could have calculated the option premium more simply by finding the Q probabilities. It turns out that the answer is yes. She could have inferred the Q probabilities from the market price of GBP-USD spot and used these to correctly price the option. Applying Eq. (1.2) to the spot price reveals that Q(Hike) = Q(Hold) = 0.5, since these values solve

$$
1.30 = \text{Q(Hike)} \times 1.31 + \text{Q(Hold)} \times 1.29. \tag{1.4}
$$

It is no coincidence that the Q probabilities that price GBP-USD spot are the same ones that price the option. Indeed, this example exhibits the fundamental theorem of asset pricing, which says that arbitrage is not possible if and only if there exist Q probabilities to price all assets, including their derivatives.

We have therefore confirmed that the absence of arbitrage ensures that investors use the same Q probabilities to price spot and derivatives, and that the risk premium associated with both are consistent.

Finally, it is interesting to consider the case where the trader priced the option using the objective probabilities of P(Hike) = 0.8 and P(Hold) = 0.2. She would have calculated the price using Eq. (1.1) as

$$
\begin{aligned}
V_t^P &= 100\text{m} \times \begin{bmatrix} \text{P(Hike)} \times \max(1.31 - 1.30, 0) + \text{P(Hold)} \\ \times \max(1.29 - 1.30, 0) \end{bmatrix} \\
&= 100\text{m} \times [0.8 \times 0.01 + 0.2 \times 0] \\
&= 800 \text{ thousand USD.}
\end{aligned}
$$

However, this price would leave the market maker open to arbitrage. An investor better versed in the theory of risk premiums would simply sell the option to the market maker at 800 thousand USD, and buy 50 million

GBP worth of GBP-USD. The investor's liability in both the hike and hold scenarios is 500 thousand USD, and therefore the investor is able to earn an arbitrage profit of 300 thousand USD.

1.2 Microeconomic Foundations of Risk Premiums

The three examples presented in this chapter provided a whirlwind tour of the concept of a risk premium and how it may arise. This section proceeds into the microeconomic foundations of the idea. I discuss the concept of utility functions to model investor preferences, and how the concavity of this utility function leads to a desire to smooth consumption, and subsequently to risk premiums. I also show how the utility function links together the P and Q probabilities.

1.2.1 Utility Functions

Economists model investor preferences using a utility function defined over consumption, $u(c_t)$, where c_t denotes the level of consumption. Utility functions should satisfy the following two requirements. First, investors prefer more consumption to less, and so utility should increase with higher levels of consumption, $u'(c_t) > 0$.[4] Second, it is reasonable to assume that the first unit of consumption is more valuable than the second, the second more valuable than the third, and so on. Each additional bite of an ice cream remains satisfying, but less so than the previous one. This is called diminishing marginal utility and it is modelled as $u''(c_t) \leq 0$.

I provide some functional forms for the utility function later in the chapter, but at this point I ask the reader to note just one important point: When consumption is low, marginal utility, $u'(c_t)$, is high, and when consumption is high, marginal utility is low. We will see that it is this property that drives risk premiums. Figure 1.3 provides an example of a utility function that satisfies the requirements mentioned in the previous paragraph, and illustrates this point.

[4] I use the $'$ notation to denote the partial derivative: $u'(c_t) \equiv \frac{\partial u(c_t)}{\partial c_t}$ and $u''(c_t) \equiv \frac{\partial^2 u(c_t)}{\partial c_t^2}$. Appendix C.1 provides a refresher on partial derivatives.

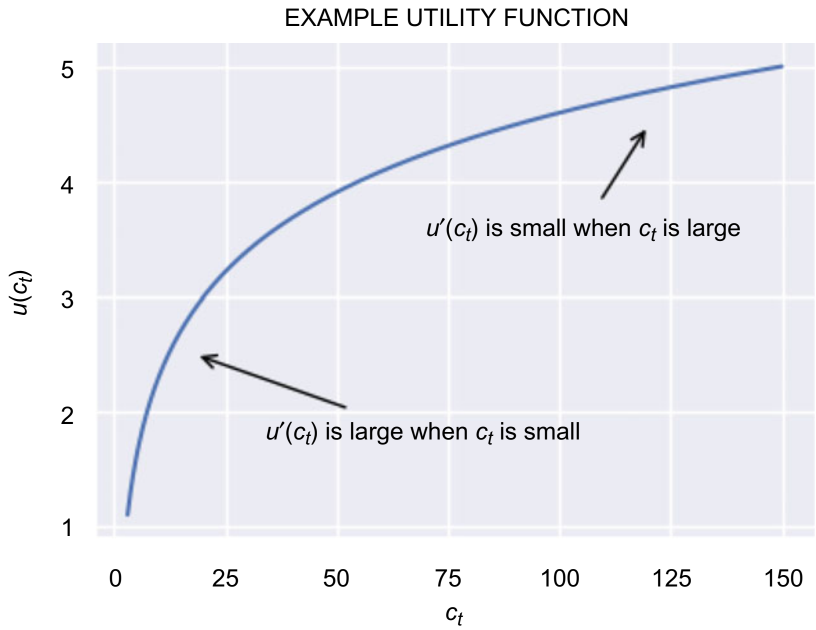


Fig. 1.3 The figure shows an example utility function to model investor preferences. Note that $u'(c_t) > 0$ means that investors prefer more consumption to less. Note also that diminishing marginal utility, $u''(c_t) < 0$, means that $u'(c_t)$ is large when c_t is small and $u'(c_t)$ is small when c_t is large

1.2.2 A One-Period Model

The investor's lifetime utility in a one-period model is

$$U(c_t, c_{t+1}) = u(c_t) + \beta \mathbb{E}_t^P[u(c_{t+1})], \tag{1.5}$$

where $u(c_t)$ denotes the utility of today's consumption and $\mathbb{E}_t^P[u(c_{t+1})]$ is the expected utility in the next period, based on consumption c_{t+1}. β is a time preference parameter that influences the degree to which investors value consumption today over consumption in the future. I show in Sect. 1.6.2 that β is closely tied to the real interest rate.

Asset markets provide the means for investors to shift consumption from today into the future. They face a trade-off. They can consume more today, and buy fewer assets and therefore expect to consume less in the next period, or vice versa. The degree to which they do so depends on the payoffs that asset markets provide, and the degree to which assets allow investors to smooth their consumption over time. Let us express this idea more formally.

Let V_t, denote the price of the asset and let X_{t+1} be its random payoff in the next period. Investors solve the following problem. They choose the number of assets n to buy so as to maximise their lifetime utility as given in Eq. (1.5). That is, the investor solves

$$\max_n (u(c_t) + \beta \mathbb{E}_t^P[u(c_{t+1})]) \tag{1.6}$$

such that

$$c_t = e_t - V_t n \tag{1.7}$$

$$c_{t+1} = e_{t+1} + nX_{t+1}, \tag{1.8}$$

where e_t and e_{t+1} denote the investor's endowments in the current and next period, respectively. I assume that e_t and e_{t+1} are exogenously set. The intuition behind the constraints in Eqs. (1.7) and (1.8) is straightforward. If the investor purchases $n = 0$ assets, then her consumption at times t and $t + 1$ would simply be her endowments; $c_t = e_t$ and $c_{t+1} = e_{t+1}$.

Solving the optimisation problem in Eq. (1.6) is straightforward. Substitute Eqs. (1.7) and (1.8) into Eq. (1.6) and set the derivative to zero as follows,

$$\frac{\partial}{\partial n}(u(e_t - V_t n) + \beta \mathbb{E}_t^P[u(e_{t+1} + nX_{t+1}]) = 0.$$

This leads to the central pricing equation,

$$V_t u'(c_t) = \mathbb{E}_t^P[\beta u'(c_{t+1}) X_{t+1}] \tag{1.9}$$

The intuition behind Eq. (1.9) can be understood as follows. If the investor bought one further unit of the asset, the left-hand side is, to first order, her drop in utility due to the smaller amount that she may consume today. This equals her expected gain in utility in the next period from having the asset's payoff X_{t+1}, appropriately discounted with her time preferences, β. This is the right-hand side.

Equation (1.9) is commonly expressed in re-arranged form as,

$$V_t = \mathbb{E}_t^P\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}X_{t+1}\right]. \tag{1.10}$$

This is arguably the most useful equation in asset pricing and I suggest that the reader commits this to memory. Let us dive deeper into its implications.

1.2.3 Exploring the Central Pricing Equation

The main point to note from Eq. (1.10) is that the correlation of the asset's payoff with marginal utility matters. If X_{t+1} is high in states of the economy where c_{t+1} is high then this asset gets a lower price. The reason is that when c_{t+1} is high, $u'(c_{t+1})$ is low. The model discounts payoffs that occur when the investor is able to consume well and perform badly when an investor's consumption is low. Such assets add risk to an investor's portfolio due to this positive correlation. The other side of the same coin is that assets that pay off well in states of the economy where c_{t+1} is low, and therefore $u'(c_{t+1})$ is high get a higher price because such assets act as a hedge. Put another way, investors bid up assets that allow them to smooth consumption relative to those that add volatility to their consumption. The remainder of this subsection expresses this idea more formally.

First, rewrite Eq. (1.10) as

$$\begin{aligned} V_t &= \mathbb{E}_t^P\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\right]\mathbb{E}_t^P[X_{t+1}] + \text{cov}_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right) \\ &= \mathbb{E}_t^P\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\right]\mathbb{E}_t^P[X_{t+1}] \\ &\quad + \rho_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right)\sigma_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}\right)\sigma_t(X_{t+1}), \end{aligned} \tag{1.11}$$

where $\text{cov}(X, Y)$ denotes the covariance between random variables X and Y, $\rho(X, Y)$ denotes their correlation, and $\sigma(X)$ denotes the volatility of X. The first line in Eq. (1.11) follows from the definition of the covariance, and the second line follows from the first, and the definition of correlation in terms of covariance. The t subscripts denote that these quantities are calculated conditional on the information available at time t.

Equation (1.10) applies to all assets. It must therefore price the risk-free zero-coupon bond. Such a bond pays 1 at maturity. I continue to assume that interest rates are zero at this stage. Therefore, the bond price today is 1.

Since the bond is risk-free, $X_{t+1} = 1$ in all states of the economy and

$$1 = \mathbb{E}_t^P\left[\beta \frac{u'(c_{t+1})}{u'(c_t)}\right]. \tag{1.12}$$

Finally, substitute Eq. (1.12) into the second line of Eq. (1.11) to reveal that

$$\begin{aligned} V_t &= \mathbb{E}_t^P[X_{t+1}] + \rho_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right)\sigma_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}\right)\sigma_t(X_{t+1}), \\ &= V_t^P + \underbrace{\rho_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right)}_{\text{1. Correlation.}} \underbrace{\sigma_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}\right)}_{\text{2. Vol. of marginal utility.}} \underbrace{\sigma_t(X_{t+1})}_{\text{3. Asset Vol.}}, \end{aligned} \tag{1.13}$$

where in the last line I have used the definition of the actuarial value $V_t^P \equiv \mathbb{E}_t^P[X_{t+1}]$. In words, this equation says that the market price of any asset is equal to its actuarial value, plus a risk premium. The risk premium depends on three things.

The first is the correlation between the payoff and marginal utility, $\rho_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right)$. If this correlation is positive, meaning that the asset tends to pay off when marginal utility is high (i.e. when consumption is low), then the price of the asset is higher than its actuarial value.

The second is the volatility of marginal utility itself, $\sigma_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}\right)$. The direction of impact of this term on the total risk premium depends on the sign of term 1. If consumption is volatile, then investors require a larger risk premium to hold assets that exhibit positive correlation with consumption (term 1 is negative). This reflects investors' desire to smooth consumption and avoid consumption volatility. Consumption is likely most volatile during times of economic turmoil.

The third is the volatility of the asset payoff itself, $\sigma_t(X_{t+1})$. The product of the first and second terms may be considered the premium per unit of risk associated with this particular asset. The third term then denotes the amount of risk.

With Eq. (1.13) in place, the next step is to build further intuition by exploring two interesting special cases of Eq. (1.13). The first considers the pricing of a bet on a coin toss versus that of a political election. Since the risk associated with the coin toss is idiosyncratic, I show that the price of the coin toss is equal to its actuarial value.

The second considers so-called *risk-neutral* investors. I show that a linear utility function leads to assets being priced at their actuarial value.

1.2.4 Special Case 1: Pricing Coin Tosses and Elections

A coin toss that pays \$1 if the coin lands on heads and \$0 if it lands on tails has $V_t^P = \$0.5$. However, since such a coin toss is unlikely to be correlated to the investor's consumption, term 1 in Eq. (1.13) is 0. Therefore, for the coin toss, $V_t = V_t^P = \$0.5$. That is, the bet on the coin trades at its actuarially fair price.

Contrast this to a bet that pays \$1 if EDP win the election and \$0 otherwise from the example in Sect. 1.1.2. Like the coin, the actuarial value of the bet is $V_t^P = \$0.5$. However, term 1 in Eq. (1.13) is now positive. If EDP wins the election, the investor's consumption will be low, and therefore marginal utility will be high, and the bet will pay off. Term 2 is positive because, by assumption, $u'(c_t)$ is positive. Finally, term 3 is a volatility, and it is therefore always positive. Therefore, the price of the bet on EDP is $V_t > V_t^P = \$0.5$. The bet on EDP trades at a premium to the bet on the coin, even though they both have the same actuarial P probabilities. This premium is driven by investor desire to smooth consumption over time.

1.2.5 Special Case 2: Risk Neutral Investors

Note that if $u'(c) = \alpha$, where α is a constant, then Eq. (1.13) reduces to

$$V_t = V_t^P$$

because terms 1 and 2 are zero when marginal utility is constant. Assets are priced at their actuarial fair value and there is no risk premium. For this reason, an investor with a utility function that is not curved, but rather linear in her consumption is called *risk-neutral*. This subsection re-enforces the idea that risk premiums are driven by curvature in the utility function. I return to this topic in Sect. 1.6.1.

1.3 Idiosyncratic Risk

The special case of a coin toss in Sect. 1.2.4 showed that when the risk associated with a payoff is entirely idiosyncratic, then it earns a zero-risk premium and the payoff is priced at its actuarial value. This section generalises this example to show that the component of the asset's payoff that is idiosyncratic earns no risk premium.

Write the asset's payoff as

$$X_{t+1} = X_{t+1}^{\text{sys}} + \varepsilon_{t+1},$$

where ε_{t+1} is uncorrelated with consumption (and thereby the wider economy), and $\mathbb{E}_t^P[\varepsilon_{t+1}] = 0$. X_{t+1}^{sys} is the *systematic* component of the asset's payoff, meaning that it is the component that is correlated with consumption. Substituting this decomposition into term 1 of Eq. (1.13) gives us

$$\begin{aligned}
\rho_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right) &= \frac{\text{cov}_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}^{\text{sys}} + \varepsilon_{t+1}\right)}{\sigma_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}\right)\sigma_t(X_{t+1})} \\
&= \frac{\text{cov}_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}^{\text{sys}}\right)}{\sigma_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}\right)\sigma_t(X_{t+1})} \\
&= \frac{\rho_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}^{\text{sys}}\right)\sigma_t\left(X_{t+1}^{\text{sys}}\right)}{\sigma_t(X_{t+1})}.
\end{aligned} \tag{1.14}$$

Finally, substitute Eq. (1.14) into Eq. (1.13) to find that the price of payoff X_{t+1} is,

$$V_t = V_t^P + \rho_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}^{\text{sys}}\right)\sigma_t\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}\right)\sigma_t\left(X_{t+1}^{\text{sys}}\right).$$

In words, the price of an asset with payoff X_{t+1} is the same as an asset with payoff X_{t+1}^{sys}. The additional volatility ε_{t+1} that payoff X_{t+1} exhibits over X_{t+1}^{sys} is not compensated with a risk premium. The important take away is the following. Investors are not compensated for all risk and assets that exhibit higher volatility do not necessarily earn investors a higher risk premium. What is important to investors is risk that is correlated with the wider economy, because this affects their consumption. Only those risks are compensated with a risk premium.

1.4 Linking *P* and *Q* Probabilities

Let us explore the link between the objective P and risk-adjusted Q probabilities. Consider a simple discrete state model of the economy. Expanding Eq. (1.10) we have

$$
\begin{aligned}
V_t &= \mathbb{E}_t^P\left[\beta \frac{u'(c_{t+1})}{u'(c_t)} X_{t+1}\right] \\
&= \sum_s \beta \frac{u'(c_{t+1}(s))}{u'(c_t)} X_{t+1}(s) \times \mathrm{P(s)},
\end{aligned} \tag{1.15}
$$

where $P(s)$ is the probability of state s occurring and $X_{t+1}(s)$ and $u'(c_{t+1}(s))$ are the payoff and marginal utility in state s, respectively. Similarly expanding Eq. (1.2) reveals

$$
V_t = \sum_s X_{t+1}(s) \times \mathrm{Q(s)}.
$$

Therefore,

$$
\frac{\mathrm{Q(s)}}{\mathrm{P(s)}} = \beta \frac{u'(c_{t+1}(s))}{u'(c_t)}. \tag{1.16}
$$

In words, the probabilities used to price assets, Q(s), are the objective probabilities, P(s), scaled by marginal utility. Since marginal utility is high when consumption is low, the investor assigns a greater probability weight than the objective probability weight in states of the world when consumption is low (economic contractions) than states of the world when consumption is high (economic booms).

Much academic literature refers to the Q probabilities as *risk-neutral* probabilities rather than risk-adjusted probabilities. The reasoning is that, once the Q probabilities are known, then assets are priced *as if* investors are risk-neutral in that prices are Q expected payoffs. However, Eq. (1.16) justifies the nomenclature of the Q probabilities as risk-adjusted probabilities.

The (optional) next section links the discussion in this chapter to asset and derivative pricing theory in continuous time. However, it should be noted that it is aimed at readers with at least some prior formal training in continuous time financial mathematics. The remainder of the chapter does not depend on knowledge from the feature box and the reader may choose to skip to the next section.

1.5 Risk Premiums in Continuous Time Models (Optional)

This section is aimed at readers with prior exposure to the dynamic hedging approach to derivatives pricing typically taken in financial mathematics textbooks and courses. I assume familiarity with the important ideas from stochastic calculus and option pricing theory.[5] The broad aim is to link the asset pricing equations from dynamic hedging with the central pricing equation developed in this chapter and to demonstrate to the reader the consistency between the two approaches.

1.5.1 Deriving *P* and *Q* Probabilities Through Dynamic Hedging

To begin, substitute Eq. (1.16) into Eq. (1.15) to find that the central pricing equation can be written as,

$$\begin{aligned} V_t &= \sum_s \frac{\mathrm{Q}(s)}{\mathrm{P}(s)} X_{t+1}(s) \times \mathrm{P}(s), \\ &= \mathbb{E}_t^P \left[\frac{\mathrm{Q}}{\mathrm{P}} X_{t+1} \right], \end{aligned} \tag{1.17}$$

where Q and P can be thought of like random variables that take the values $\mathrm{Q}(s)$ and $\mathrm{P}(s)$ in state s of the economy at time $t+1$. Our first task is to derive a continuous time analogue of this equation.

Let us begin, as Black and Scholes (1973) and Merton (1974) did in their seminal work on option pricing, and assume that the price of an asset, S_t, follows a log-normal stochastic process,

$$\frac{\mathrm{d}S_t}{S_t} = \mu \mathrm{d}t + \sigma \mathrm{d}Z_t^P, \tag{1.18}$$

[5] Interested readers without a background in these topics may consult Iqbal (2018) for a full treatment of option pricing and Shreve (2000) for a full treatment of continuous time financial mathematics.

where μ is the (constant) expected return of the asset, σ is its (constant) volatility and Z_t^P is a Brownian motion under the objective P measure. The reader should think of this equation as empirically motivated.[6]

The dynamic hedging approach to price a derivative on S_t proceeds in three steps. First, form a portfolio of long the derivative, V_t and short the Δ_t units of the asset. The value of this portfolio is,

$$W_t = V_t - \Delta_t S_t.$$

The dynamics of the portfolio are then

$$\begin{aligned} \mathrm{d}W_t =& \mathrm{d}V_t - \Delta_t \mathrm{d}S_t \\ =& \underbrace{\frac{\partial V(S_t,t)}{\partial t}\mathrm{d}t}_{\text{1. Exposure to Time}} + \underbrace{\left(\frac{\partial V(S_t,t)}{\partial S} - \Delta_t\right)\mathrm{d}S_t}_{\text{2. Exposure to Spot Moving}} + \underbrace{\frac{1}{2}\frac{\partial^2 V(S_t,t)}{\partial S^2}\mathrm{d}S_t^2}_{\text{3. Gamma Term}} \end{aligned} \tag{1.19}$$

Second, execute a dynamic delta hedging strategy. A delta hedging trader holds $\Delta_t = \frac{\partial V(S_t,t)}{\partial S}$ to remove any exposure to spot moving. This sets term 2 to zero. Also, from (1.18) we have that $\mathrm{d}S_t^2 = \sigma^2 S_t^2 \mathrm{d}t$. Therefore Eq. (1.19) becomes,

$$\mathrm{d}W_t = \left(\frac{\partial V(S_t,t)}{\partial t} + \frac{1}{2}\frac{\partial^2 V(S_t,t)}{\partial S^2}\sigma^2 S_t^2\right)\mathrm{d}t. \tag{1.20}$$

With the delta hedge in place, all risk in the portfolio has been removed in that there is no $\mathrm{d}Z_t^P$ term in Eq. (1.20).

Third, assume that arbitrage is impossible. Since the portfolio is risk-free, its value must grow at the risk-free rate, which is zero under the assumption so far in this chapter that interest rates are zero. Therefore, $\mathrm{d}W_t = 0$ and we have the Black-Scholes-Merton (BSM) partial differential equation (PDE):

$$\frac{\partial V(S_t,t)}{\partial t} = -\frac{1}{2}\frac{\partial^2 V(S_t,t)}{\partial S^2}\sigma^2 S_t^2, \tag{1.21}$$

with the boundary condition being the payoff of the derivative, $\psi(S_T)$. For example, in the case of a vanilla call option, $\psi(S_T) = \max(S_T - K, 0)$. The

---

[6] I do not suggest that this equation describes real market dynamics. Indeed, skewness and kurtosis are well-documented features of asset price returns that are not modelled by this stochastic process. I simply start with arguably the simplest *empirical* model to illustrate the key ideas.

Feynman-Kac theorem then tells us that the solution to

$$V_t(S_t, t) = \mathbb{E}_t^Q[\psi(S_T))], \tag{1.22}$$

with

$$\frac{dS_t}{S_t} = \sigma\, dZ_t^Q, \tag{1.23}$$

where Z_t^Q is a Brownian motion under the Q measure solves the PDE in Eq. (1.21). To see this, simply calculate the dynamics of the martingale in Eq. (1.22) and set the co-efficient of $dt = 0$. I leave this to the reader as an exercise.

Comparing Eqs. (1.23) and (1.18) we see that

$$dZ_t^Q = dZ_t^P + \frac{\mu}{\sigma}dt. \tag{1.24}$$

Therefore, via Girsanov's theorem, we can write the asset price as

$$S_t = \mathbb{E}_t^P\left[\frac{\xi_T}{\xi_t}S_T\right], \tag{1.25}$$

where

$$\xi_s = \exp\left(-\frac{\mu}{\sigma}Z_s^P - \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 s\right).$$

We can quickly check that Eq. (1.25) works as follows. The integrated form of Eq. (1.18) is,

$$S_T = S_t \exp\left(\mu(T-t) - \frac{1}{2}\sigma^2(T-t) + \sigma(W_T - W_t)\right).$$

Next, calculate the expectation together with $\frac{\xi_T}{\xi_t}$ as below to find that

$$\begin{aligned}\mathbb{E}_t^P\left[\frac{\xi_T}{\xi_t}S_T\right] &= S_t \exp\left(\left(\mu - \frac{1}{2}\sigma^2 - \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2\right)(T-t)\right) \\ &\quad \times \mathbb{E}_t^P\left[\exp\left(\sigma - \frac{\mu}{\sigma}\right)(W_T - W_t)\right] \\ &= S_t,\end{aligned}$$

as required.[7] Finally, note that ξ_t is a martingale and $\xi_0 = 1$. Therefore, $\mathbb{E}^P[\xi_T] = 1$ and we can define an equivalent probability measure Q by

$$\frac{dQ}{dP} = \xi_T.$$

We have already shown that Q prices assets and we can write,

$$\xi_t S_t = \mathbb{E}_t^P\left[\frac{dQ}{dP} S_T\right]. \tag{1.26}$$

The Radon–Nikodym derivative, $\frac{dQ}{dP}$, is analogous to $\frac{\mathrm{Q}}{\mathrm{P}}$ in Eq. (1.17) and Eq. (1.17) is analogous to Eq. (1.26).

We have shown that asset pricing equations that result from dynamic hedging arguments are consistent in their form with our central pricing equation derived from utility maximisation in that assets are priced using *risk-adjusted* Q probabilities rather than objective P probabilities. What remains is to relate $\xi_T = \frac{dQ}{dP}$ to marginal utilities in the way that $\frac{\mathrm{Q}}{\mathrm{P}}$ related to marginal utility in Eq. (1.16). This is the topic of the next subsection.

1.5.2 The Central Pricing Equation in Continuous Time

The investor must choose the number of assets, n, to purchase at time t. She arrives at the following first-order condition,

$$S_t u'(c_t) = \mathbb{E}_t[e^{-\delta\varepsilon} u'(c_{t+\varepsilon}) S_{t+\varepsilon}], \tag{1.27}$$

where δ determines her time preference, and ε is *small* period of time. This equation can be understood in a manner similar to its discrete time counterpart that we derived in Sect. 1.2.2.

If the investor were to purchase n units of the security, her consumption reduces from $e_t dt$ to $c_t dt = e_t dt - nS_t$. Her utility loss today is $u'(c_t)(e_t - c_t)dt = u'(c_t)nS_t$. This is the left-hand-side of Eq. (1.27). Having bought n securities, her consumption at time $t+\varepsilon$ is $c_{t+\varepsilon}dt = e_{t+\varepsilon}dt + nS_t$. Her expected utility gain is therefore $\mathbb{E}_t[e^{-\delta\varepsilon}u'(c_{t+\varepsilon})(c_{t+\varepsilon} - e_{t+\varepsilon})dt] = \mathbb{E}_t[e^{-\delta\varepsilon}u'(c_{t+\varepsilon})S_{t+\varepsilon}]$, which is the right-hand-side of Eq. (1.27).

[7] Recall that $\mathbb{E}[\exp(X)] = \exp(\mathbb{E}[X] + \frac{1}{2}\mathrm{var}[X])$ for a normally distributed random variable X.

Finally rewrite Eq. (1.27) as

$$S_t u'(c_t) e^{-\delta t} = \mathbb{E}_t[e^{-\delta(t+\varepsilon)} u'(c_{t+\varepsilon}) S_{t+\varepsilon}], \tag{1.28}$$

and note that this equation is identical to Eq. (1.25) with ξ_t set to $\xi_t = e^{-\delta t} u'(c_t)$. We therefore have a continuous time analogue of our central pricing equation where the Q and P probabilities are related through marginal utility.

1.6 Interest Rates

In this section I relax the assumption of zero interest rates and apply the central pricing equation to understand the behaviour of the real risk-free interest rate. I show that the central pricing equation predicts a strong bid for safe assets in times of economic turmoil from risk-averse investors. First, investors' lower expectations for future growth and higher uncertainty over their future consumption streams leads to a demand for saving. The first feature box illustrates this idea using consumer credit data over the Coronavirus recession of 2020, and the Global Financial Crisis (GFC) from the United Kingdom (U.K.). Next, this higher demand for savings leads to lower real interest rates. The second feature box assesses this formulation in the context of the behaviour of U.S. real government bond yields over these two past recessions.

To pin down the ideas, I choose a functional form for investor utility functions, namely the constant relative risk aversion (CRRA) utility.

1.6.1 CRRA Utility

There are many forms that a utility function can take. However, for our purposes the so-called (CRRA) utility function is sufficient to demonstrate the main ideas. CRRA utility is written as,

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \tag{1.29}$$

where γ determines the degree of risk aversion. $\gamma = 0$ denotes a risk-neutral investor, because $u'(c_t) = 1$ if $\gamma = 0$. The investor becomes more risk-averse as γ increases. The $\gamma = 1$ case corresponds to the log utility function from Fig. 1.3.

The CRRA utility function gets its name from the property that it has a constant response to a *relative* change in consumption. A 10% fall in consumption leads to a proportional fall in utility. To see this, note that

$$\ln\left(\frac{u(c)}{u(c_0)}\right) = (1-\gamma)\ln\left(\frac{c}{c_0}\right).$$

Let us now apply the CRRA utility and the central pricing equation to calculate the risk free interest rate.

1.6.2 The Real Risk-Free Interest Rate

The continuously compounded real risk-free interest rate, r_t^{real}, is defined via

$$Q_t^{\text{real}} \equiv \exp(-r_t^{\text{real}}), \tag{1.30}$$

where Q_t^{real} is the real price of the risk-free zero-coupon inflation linked bond, paying one unit of real output at maturity.[8] Applying Eq. (1.10) to price the bond gives us

$$Q_t^{\text{real}} = \mathbb{E}_t^P\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\right] \tag{1.31}$$

$$= \mathbb{E}_t^P\left[\beta\left(\frac{c_t+1}{c_t}\right)^{-\gamma}\right], \tag{1.32}$$

where, in the second line, I substitute in the derivative of Eq. (1.29). To gain further intuition, I follow the approach of Cochrane (2005) and assume that consumption growth is log-normally distributed. That is,

$$\ln\frac{c_{t+1}}{c_t} \sim \mathcal{N}(\mu,\sigma^2),$$

where $\mu = \mathbb{E}_t^P\left[\ln\frac{c_t+1}{c_t}\right]$ and $\sigma^2 = \sigma^2\left(\ln\frac{c_{t+1}}{c_t}\right)$. B_t is then calculated as

$$Q_t^{\text{real}} = \mathbb{E}_t^P\left[e^{\left(\ln\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right)\right)}\right]$$

[8] I discuss the relationship between real and nominal variables in detail over the following chapters.

$$= e^{\ln\beta}\mathbb{E}_t^P\left[e^{-\gamma\ln\left(\frac{c_{t+1}}{c_t}\right)}\right]$$
$$= e^{\ln\beta - \gamma\mu + \frac{1}{2}\gamma^2\sigma^2}. \tag{1.33}$$

where the final line follows from the standard result that $\mathbb{E}[\exp(X)] = \exp(\mathbb{E}[X] + \frac{1}{2}\text{var}[X])$ for a normal random variable X. Finally, taking logs of both sides of Eq. (1.33) and expressing it in terms of r_t^{real} using Eq. (1.30) we have,

$$r_t^{\text{real}} = \underbrace{-\ln\beta}_{\text{1. Impatience.}} + \underbrace{\gamma\mathbb{E}_t^P\left[\ln\frac{c_{t+1}}{c_t}\right]}_{\text{2. Growth Expectations.}} - \underbrace{\frac{1}{2}\gamma^2\sigma_t^2\left(\ln\frac{c_{t+1}}{c_t}\right)}_{\text{3. Precautionary Savings.}}. \tag{1.34}$$

Equation (1.34) is key to understanding movements in real interest rates.

The *impatience* term reflects investors' time preferences. If β is small, then investors value future consumption less, and want to consume today. Small β leads to a higher value of $-\ln\beta$ and hence a higher real interest rate. The higher interest rate is required to convince impatient investors to save today.[9]

The *growth expectations* term leads to lower interest rates when expectations of consumption growth, $\mathbb{E}_t^P\left[\ln\frac{c_{t+1}}{c_t}\right]$, are weak. In an economy that may enter recession, the investor needs to save more today to avoid a shortfall in consumption in the next period, and this higher savings rate drives interest rates lower. Note that γ appears here. The greater the level of risk aversion among investors, the greater the impact of this term on interest rates.

Finally, *precautionary savings* lowers the interest rate. Its impact is large when economic uncertainty is high, because this is when volatility of consumption growth, $\sigma_t^2\left(\ln\frac{c_{t+1}}{c_t}\right)$, is high. Since this term is multiplied by γ^2, again, the impact of precautionary savings on the interest rate increases with higher investor risk aversion.

The next feature box discusses Eq. (1.34) in the context of the recession associated with the GFC from late 2007 to 2009, and in the context of the recession driven by the Coronavirus pandemic in 2020.

[9] We will see in Sect. 7.6.1 that $-\ln\beta$ is closely related to the commonly cited r^*, natural rate of interest is expected to prevail when an economy is at full strength and inflation is stable.

Consumer Behaviour and the State of the Economy

Figure 1.4 plots household credit data from the U.K. during past recessions. The Coronavirus recession reached its depths during Q1 and Q2 of 2020. The upper plot shows that consumers repaid loans during this period of economic turmoil. The sudden nature of the recession caused a rapid change in consumer behaviour.

The lower plot shows the GFC. The recession ran in the U.K. for five consecutive quarters beginning in 2008. The GFC unfolded more slowly than the Coronavirus recession and consumers appear to have adjusted their behaviour more slowly also, with a slow higher in consumer saving beginning in Q1 of 2008.

There may be other explanations for these trends. For example, lenders may tighten lending standards, and widen credit spreads to increase the cost of credit to consumers during recessions and this may drive consumers to borrow less. However, the data is at least consistent with the intuition that we developed in Sect. 1.6.2, where we postulated that lower growth expectations and higher precautionary savings during recessions drive consumers to save rather than consume.

The Real Interest Rate and the State of the Economy

Figure 1.5 plots 5-year US real and nominal interest rates over the GFC and over the Coronavirus recession. How does Eq. (1.34) do in explaining the observed behaviour?

As the economic crises hit, three changes are likely to have occurred. Investors (i) lowered their expectations of consumption growth, (ii) raised their estimate of forward-looking consumption volatility as they became more uncertain over the path of the economy and (iii) are likely to have become more fearful and risk-averse, and therefore γ may have increased. All three changes predict a fall in the real interest rate.

This fall is broadly what is observed in the data. Both nominal and real interest rates fell somewhat dramatically during the crisis periods. However, there are several phenomena in the data that are not well explained in our model.

First, there was a spike in real interest rates in late 2008. This is surprising in the context of the model. It is unlikely that expectations of growth had recovered, because US equity markets

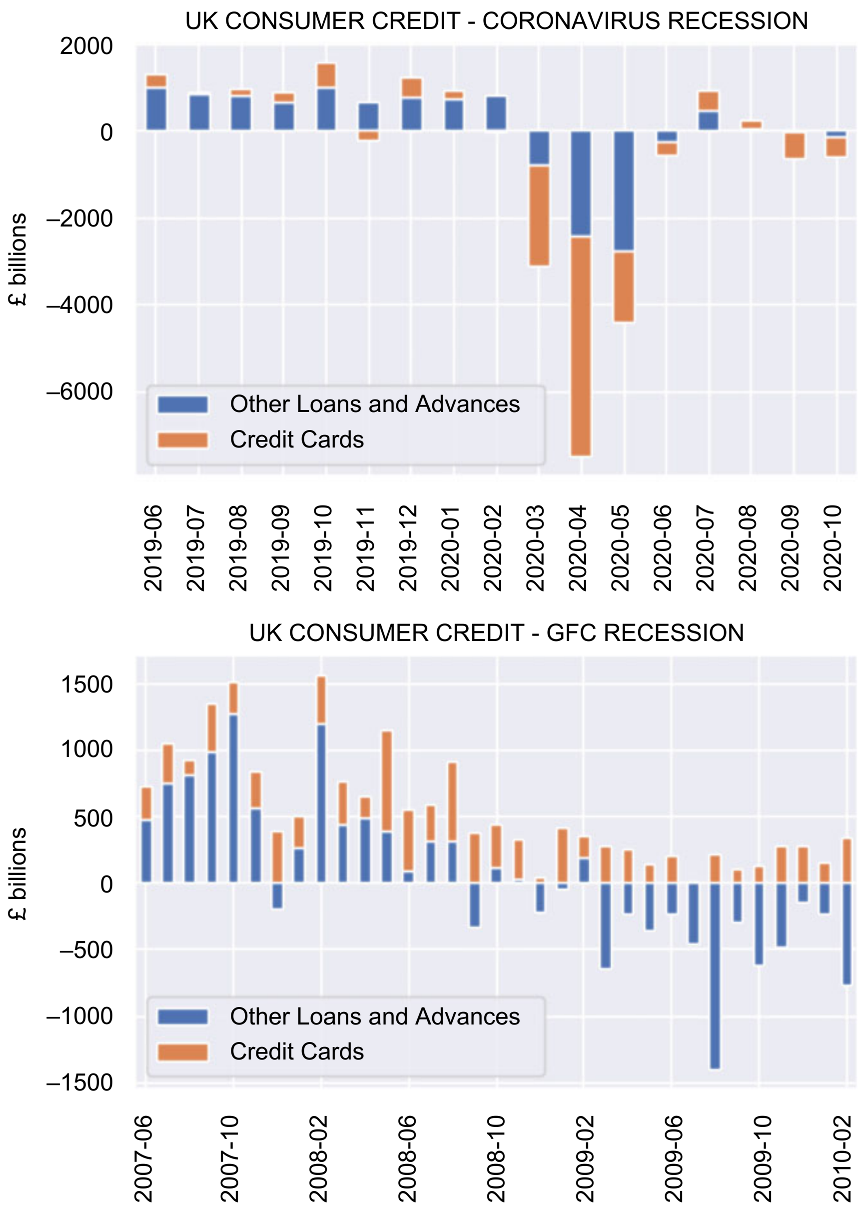


Fig. 1.4 The chart shows the behaviour of consumers over two past recessions. Consumers moved towards saving during economic uncertainty, indicated by the downward bars. The Coronavirus recession took place during Q1 and Q2 of 2020. The GFC recession lasted over five consecutive quarters beginning in Q2 of 2008 (*Source* Bank of England, household credit data; https://www.bankofengland.co.uk/statistics/visual-summaries/household-credit, December 20, 2020)

remained far from recovery at that stage, with the S&P500 returning to its pre-crisis levels only in around 2011. Second, economic uncertainty remained high. For instance, the VIX, a measure of equity market volatility, remained at elevated levels until around mid-June 2009. However, this spike in real interest rates quickly reversed and a trend lower in real interest rates subsequently resumed. Real interest rates passed 0 and turned negative in Q4 of 2010.

Second, the reader may have expected real interest rates to recover at an earlier stage than the eventual turning point observed in the middle of 2013. However, it should be noted that, although the financial crisis in the U.S. had abated and the economy had begun its recovery, the GFC had triggered a Eurozone sovereign debt crisis. The savings demand driven by the Eurozone crisis may have continued to influence real interest rates to remain low.

Third, real interest rates rose as our equation predicts as global economies eventually recovered, and economic uncertainty, and therefore volatility of consumption growth subsided. However, neither real nor nominal interest rates returned to pre-GFC levels.

A further point to note is that the spike in the real interest rate that was observed during the GFC recession was briefly observed again in the pandemic-driven recession. However, it was shorter lived and less impactful. I suspect that such spikes continue to decrease in size during future recessions.

Finally, note that the worsening fiscal position of the government during recessions due to lower tax receipts and higher benefit claims is a secondary effect. One may expect the increase in *sovereign credit risk* to act to increase the yields on U.S. government bonds. However, at the time of writing, markets continue to demonstrate faith in the U.S. government's ability to eventually repay debt without explicit default or default via inflation. This is generally the case in developed markets, but not always true in emerging market economies. Indeed, yields on emerging market government bonds, often increase during economic turmoil, where the increased sovereign credit risk outweighs the higher desire for savings, causing government bond yields to rise. In short, in such circumstances, government bond yields are not risk-free, and so it is not correct to apply Eq. (1.34) to make predictions. I return to this topic when discussing the Fiscal Theory of the Price Level in Sect. 7.5.

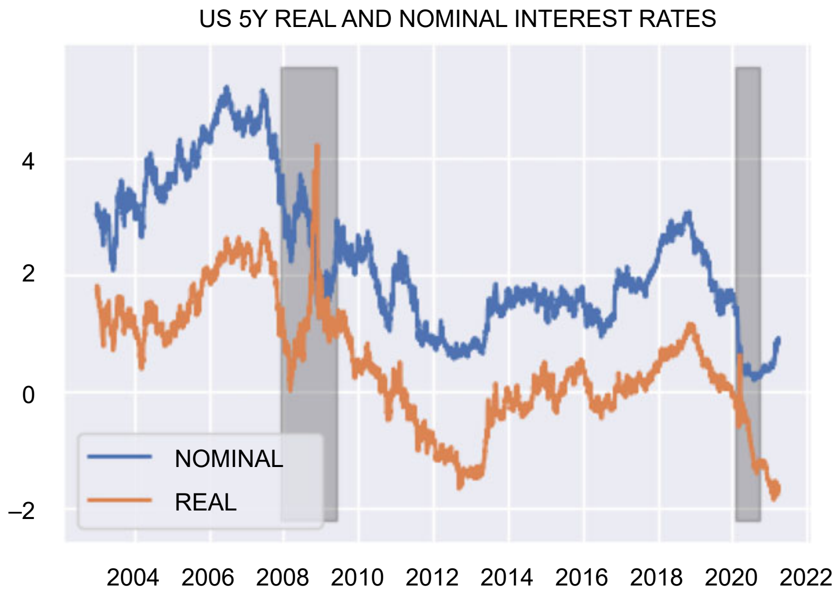


Fig. 1.5 The chart shows the behaviour of nominal and real 5Y interest rates over the previous two recessions. The real interest rate broadly fell during the recessionary periods as predicted by the central pricing equation. The feature box discusses the observed dynamics in more detail (*Source* Board of Governors of the Federal Reserve System [US], 5-Year Treasury Inflation-Indexed Security, Constant Maturity [DFII5] and 5-year Treasury Constant Maturity Rate [DGS5] retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DFII5, September 30, 2020)

1.7 Expected Returns (Optional)

Much of the discussion in finance is phrased in terms of expected returns in excess of the risk-free interest rate and so-called "betas." Indeed, practitioners often refer to *high* and *low* beta currencies and assets. The purpose of this section is to rewrite the central pricing equation expected return format. I show how to arrive at the well-known Capital Asset Pricing Model (CAPM) from the central pricing equation written in expected return form. This model likely inspired the practitioner parlance. I show that by referring to the "beta" of an asset or currency, practitioners are effectively rank ordering it in terms of its risk premium.

1.7.1 The Central Pricing Equation as an Excess Return

Let us continue from Eq. (1.11). Applying this equation to a one-period risk free bond with payoff $X_{t+1} = 1$ in all states gives us,

$$\frac{1}{R_F} \equiv Q_t^{\text{real}} = \mathbb{E}_t^P\left[\beta \frac{u'(c_{t+1})}{u'(c_t)}\right],$$

where R_F denotes the real risk free gross return. Substituting R_F back into Eq. (1.11) we find that,

$$V_t = \frac{\mathbb{E}_t^P\left[X_{t+1}\right]}{R_F} + \text{cov}_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, X_{t+1}\right). \tag{1.35}$$

Finally, let $R_{t+1} \equiv X_{t+1}/V_t$ denote the asset's gross (random) return. Dividing Eq. (1.35) through by V_t and rearranging provides the expected return relationship,

$$\mathbb{E}_t^P[R_{t+1}] - R_F = -\text{cov}_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, R_{t+1}\right). \tag{1.36}$$

This equation is interpreted analogously to Eq. (1.11). If the asset's return is high in states of the world where consumption is low (and therefore marginal utility is high), then the right-hand side of the equation above is positive, and such assets are priced with positive expected returns. The positive expected returns compensate investors for adding systematic risk to their portfolios. The converse is true for returns that are negatively correlated with consumption. Finally, note that uncorrelated, or idiosyncratic returns have an expected return of R_F, as before.

1.7.2 The CAPM

Consider the following form for marginal utility

$$\beta \frac{u'(c_{t+1})}{u'(c_t)} = -bR_{t+1}^M, \tag{1.37}$$

where b is a constant and R_{t+1}^M is the return of the market portfolio, defined as a portfolio consisting of all available securities where the proportion invested in each security corresponds to its relative market value. The relative market value of a security is equal to the aggregate market value of the security divided by the sum of the aggregate market values of all securities. In practical applications of the CAPM the market portfolio is proxied using value-weighted equity indices, such as the S&P 500.

To see how Eq. (1.37) could be a plausible form for marginal utility, consider a investor with log utility, $u(c) = \ln c$. Substituting this form of utility into Eq. (1.37) gives us

$$\beta \frac{c_t}{c_{t+1}} = -bR^M \tag{1.38}$$

Finally, taking logs on both sides we have

$$\ln \frac{c_{t+1}}{c_t} = \ln \frac{b}{\beta} + \ln R^M_{t+1}. \tag{1.39}$$

Investors hold the market portfolio in aggregate. If investors consume a fraction of their wealth at each point in time, then aggregate consumption growth may indeed be proportional to market returns as in Eq. (1.39), and Eq. (1.37) provides a plausible description of marginal utility growth. This relationship is the reason that I have referred to an asset's correlation with consumption, and correlation with an investor portfolio interchangeably.

Next, substitute Eq. (1.39) into Eq. (1.36) to find,

$$\mathbb{E}^P_t[R_{t+1}] - R_F = b\mathrm{cov}\big(R^M_{t+1}, R_{t+1}\big). \tag{1.40}$$

We are now ready to derive the CAPM. First, note that the market portfolio itself is a tradeable return. Substituting $R_{t+1} = R^M_{t+1}$ into Eq. (1.40) gives us

$$\begin{aligned} \mathbb{E}^P_t[R^M_{t+1}] - R_F &= -\mathrm{cov}\big(-bR^M_{t+1}, R^M_{t+1}\big) \\ &= b\mathrm{var}(R^M_{t+1}). \end{aligned} \tag{1.41}$$

Finally, rearranging Eq. (1.41) for b and substituting back into Eq. (1.40) gives us

$$\begin{aligned} \mathbb{E}^P_t[R_{t+1}] - R_F &= \frac{\mathrm{cov}\big(-bR^M_{t+1}, R^M_{t+1}\big)}{\mathrm{var}(R^M_{t+1})} \mathbb{E}^P_t[R^M_{t+1}] - R_F \\ &= \beta^{\mathrm{CAPM}}\big(\mathbb{E}^P_t[R^M_{t+1}] - R_F\big), \end{aligned} \tag{1.42}$$

where $\beta^{\text{CAPM}} \equiv \frac{\text{cov}(-bR^M_{t+1}, R^M_{t+1})}{\text{var}(R^M_{t+1})}$. Equation (1.42) is the CAPM. It says that the expected excess return on any asset is equal to its regression beta, multiplied by the expected return on the market portfolio as a whole.

Practitioners often estimate Eq. (1.42) by running the following regression,

$$R^i_{t+1} - R_F = \beta^{\text{CAPM},i} \times (R^M_{t+1} - R_F) + \varepsilon^i_{t+1},$$

and replace R^M with a stock index, such as the S&P500. In practitioner parlance then, *high beta* assets are those that have a large β^{CAPM} and therefore a large expected return, or risk premium. In FX, these may be EM currencies, but also some DM currencies. Negative beta currencies are those that have a negative β^{CAPM} and therefore typically move in the opposite direction to the aggregate market and therefore act as hedges to investor portfolios. These are the safe-haven currencies such as CHF, JPY and USD.

1.8 Risk Premiums and the Efficient Market Hypothesis (EMH)

The EMH is sometimes incorrectly interpreted as implying that currency and other asset prices do not offer expected profits because prices are unpredictable. Such an understanding of the EMH implies that risk premiums cannot exist in an efficient market. This chapter argues that expected profits/risk premiums exist and their size depends on the covariance of the asset payoff with marginal utility growth (Eq. 1.10). The theory of risk premiums presented in this chapter and the EMH are indeed consistent with each other and can be reconciled via the following logic.

Fama (1970) wrote "a market in which prices always 'fully reflect' available information is called 'efficient'."[10] The central pricing equation, Eq. (1.10), calculates the asset price at time t, precisely by taking into account all of the information that is available at time t. The joint probability distribution of the payoff, X_{t+1}, and marginal utility growth, $u'(c_{t+1})/u(c_t)$, is calculated based on what investors know about the asset and condition of the economy at time t. Based on Fama's definition, it is therefore clear that not only is the risk premium not a violation of the EMH, but is rather clearly accommodated within it.

[10] Eugene Fama is often thought of as the father of the EMH. He was awarded the 2013 Nobel Memorial Prize in Economics in recognition of his work in this and other areas.

1.9 Chapter Summary

- Assets do not typically trade at their actuarial value. The difference between the market price of an asset and its actuarial value is its risk premium.
- Risk premiums exist because investors' objective is to smooth consumption risk over time. Investors therefore bid up assets that hedge consumption, and offer assets that add risk to their portfolios.
- The probabilities that are backed out from the traded prices of assets are known as risk-adjusted, or Q probabilities. They are linked to the objective P probabilities through marginal utility.
- The equations that are derived from investor utility optimisation are consistent with those that result no-arbitrage (continuous or discrete time) arguments.
- The risk-free interest rate depends on expected consumption growth and volatility of consumption. These dependencies increase as investors become more risk-averse.
- Lower expectations of consumption growth leads investors to save more in the risk-free asset. This acts to lower the real interest rate during recessions.
- Higher consumption volatility leads investors to smooth their consumption stream by investing more in the safe asset. Since economic volatility tends to be higher during recessions, this leads to a lower real interest rate.
- The theory of the risk premium is consistent with and rooted in the EMH.

References

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637.

Cochrane, J. H. (2005). *Asset pricing* (Rev ed.). Princeton University Press.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417.

Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.

Iqbal, A. (2018). *Volatility*. Wiley Finance Series: Practical Options Theory.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29*(2), 449.

Shreve, S. E. (2000). *Stochastic calculus for finance II: Continuous-time models*. Springer.

2

FX Forwards and the Carry Trade

I introduce two concepts that are central to understanding movements in FX markets. The first, covered interest parity (CIP), is arguably the most important concept for FX market participants to understand. It relates the forward exchange rate to the spot exchange rate through the foreign and domestic interest rate.

Second, I discuss uncovered interest parity (UIP) and the carry trade. UIP conjectures that the forward exchange rate reflects the market's expectation of the future spot exchange rate. I apply the asset pricing model of Chapter 1 to show that UIP should not hold in general and that the risk premium is the difference between the expected future spot rate and the forward exchange rate. I derive the relationship between UIP and the well known FX carry trade to show that the expected profits from the carry trade come from risk premium in currencies. Also, I distinguish between the concept of FX *carry*, which represents a type of *buffer* against adverse movements in the FX rate, and the expected profits from carry trading.

Finally, in the chapter's feature boxes I study practical examples of risk premiums applied in currencies. One feature box studies a *risky* currency, MXN, over the Coronavirus crisis in 2020. I show that the risk premium approach may be able to explain a substantial proportion of the price movement that was observed. In the following feature box I discuss the USD, and the negative risk premium/safe-haven status conferred upon it and its so-called *exorbitant privilege*.

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_2

2.1 Covered Interest Rate Parity

Forward contracts provide the basic building block of the CIP relationship. A forward contract sets the price today at which an investor will buy (or sell) a currency at a date in the future. CIP says that this price depends on the current spot price, and the interest rate differential between the currencies. The next subsection derives this result.

2.1.1 Forwards Basics

Suppose that EUR-USD is trading at a spot rate of $S_t = 1.20$ today. A German car manufacturer exports 100 million USD worth of vehicles to the U.S. These vehicles will take 1 year to sell and therefore the exporter wishes to sell USD in exchange for buying EUR in 1 year. The manufacturer can hedge her risk by trading a *forward* contract.

A forward contract is an agreement where the manufacturer agrees to purchase underlying the currency pair at an expiry date T in the future at a strike price K that is agreed upon today. At time T the payoff of the forward contract is therefore the spot rate, less the agreed upon strike, $S_T - K$. The price that the manufacturer must pay to enter this contract clearly depends on K. The lower K, the higher the cost of this forward contract. The special value of K where the cost to enter this contract today is exactly zero is denoted by $f_t(T)$.

Let $\tau \equiv T - t$ denote the time remaining until maturity. In our example, $\tau = 1$ year and the underlying currency pair is EUR-USD. We wish to determine $f_t(T)$. It turns out that $f_t(T)$ depends only on the interest rate differential paid on EUR and USD T maturity safe assets. It is given by

$$f_t(T) = S_t e^{(r_d - r_f)\tau}. \tag{2.1}$$

Here, r_f is the continuously compounded interest rate in the base currency (EUR in the case of EUR-USD) and the r_d is the continuously compounded interest rate in the numeraire currency (USD in the case of EUR-USD).[1] For example, if $r_d = 2\%$, $r_f = 0\%$ and $S_t = 1.20$, then $f_t = 1.2240$. In our example, the German car exporter would receive 100 million USD from car sales in one year, and sell these USD in exchange for 100 million/1.2240 = 81.69 million EUR. Figure 2.1 shows the forward path corresponding to this example.

[1] The notation r_d and r_f comes from financial literature that writes currency pairs as FOR-DOM. For EUR-USD, EUR is the FOR and USD is the DOM. For USD-JPY, USD is FOR and JPY is DOM, and so on. The subscripts in r_d and r_f refer to DOM and FOR respectively.

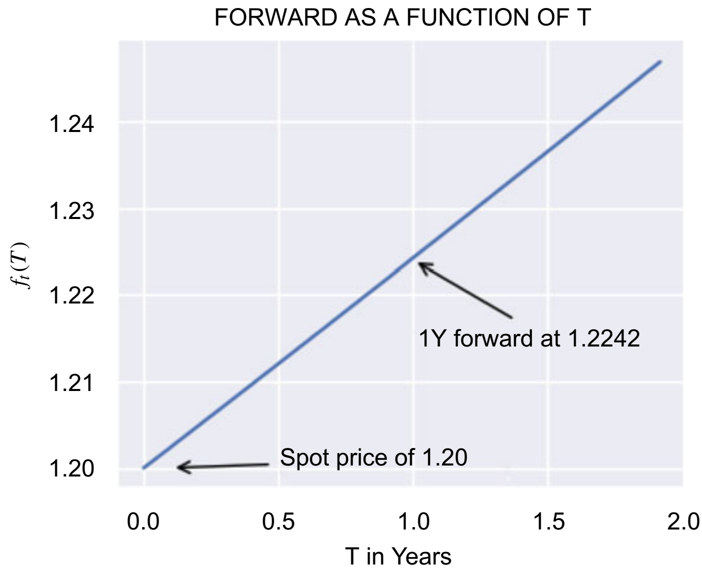


Fig. 2.1 The figure shows $f_t(T)$ as a function of T for our example. I set today to be $t = 0$. The EUR-USD spot price is 1.2020. The USD interest rate, r_d, is 2% and the EUR interest rate is 0%. This leads to the price of the 1Y forward at 1.2242. In real markets, both r_d and r_f themselves depend on T and this may lead to more complex shapes than the linear relationship

To derive Eq. (2.1), consider the following zero cost strategy. At time t, the investor borrows S_t USD to purchase 1 EUR, and invests it in a Euro-denominated safe asset at a rate of r_f. The investor simultaneously sells $e^{r_f \tau}$ EUR notional of the time T maturity forward contract at zero cost, with strike $f_t(T)$. By time T, the 1 EUR holding has grown to $e^{r_f \tau}$ EUR in value. The forward contract settles, and so the investor sells the $e^{r_f \tau}$ EUR in exchange for $e^{r_f \tau} f_t(T)$ USD. Finally, after repaying $S_t e^{r_d \tau}$ USD, which is the S_t USD originally borrowed plus interest owed, the investor's payoff is $e^{r_f \tau} f_t(T) - S_t e^{r_d \tau}$. Since this strategy cost zero to initiate, its (certain) payoff must also be zero, else there is an arbitrage profit available. Therefore, we must have $e^{r_f \tau} f_t(T) - S_t e^{r_d \tau} = 0$, which, re-arranged, is Eq. (2.1).

The next (optional) section provides a more general approach to calculating the prices of forward contracts of any strike level.

2.1.2 Calculating the Prices of Forwards (Optional)

Consider the following zero cost trading strategy. Again, I use EUR-USD as an example. First, purchase the EUR-USD forward contract with strike

K and maturity T for a price of $F_t(T) = F(S_t, K, T)$ USD. I write the price as a function, $F(S_t, K, T)$, to make explicit that the price to enter a forward contract depends on S_t, K and T. Our aim is to find K such that the contract's price is zero.

Second, hedge by borrowing a yet to be determined amount, Δ_f EUR, at a rate of r_f, sell the EUR in the market at the prevailing spot rate to receive $\Delta_f S_t$ USD. Finally, borrow an amount X USD such that the strategy costs zero. That is,

$$F(S_t, K, T) - \Delta_f S_t - X = 0. \tag{2.2}$$

At time T the forward contract pays $S_T - K$ USD. The trader must also buy back $\Delta_f e^{r_f \tau}$ EUR at price S_T to pay back the EUR she borrowed, plus the interest. She must also pay back $Xe^{r_d \tau}$ USD to return the USD she borrowed, plus the interest. The value of her portfolio at time T is therefore

$$S_T - K - \Delta_f e^{r_f \tau} S_T - Xe^{r_d \tau}.$$

If she chooses to set $\Delta_f = e^{-r_f \tau}$ ex-ante then the value of her portfolio at time T is $-K - Xe^{r_d \tau}$. This strategy cost zero to implement and carries no risk because the payoff is known at time t. The value at time T must therefore be zero to avoid an arbitrage opportunity and therefore $-K - Xe^{r_d \tau} = 0$. Solving for X and substituting back into Eq. (2.2) gives us the price of the forward contract as

$$F(S_t, K, T) = e^{-r_f \tau} S_t - Ke^{-r_d \tau}.$$

Finally, $f_t(T)$ is the level of K such that $F(S_t, T, K) = 0$. Therefore $f_t(T) = S_t e^{(r_d - r_f)\tau}$ and we can write $F_t(T) = F(S_t, K, T) = e^{-r_d \tau}(f_t(T) - K)$.

CIP in practice and the Cross-Currency Basis

Let $f_t^{\mathrm{m}}(T)$ denote the FX forward price observed in the market. The no-arbitrage argument presented in Sect. 2.1.1 suggests that $f_t^{\mathrm{m}}(T) = f_t(T)$. Indeed, this has been the case through most of FX history. However, since the GFC, there has been a gap between these two quantities. This is the so-called *cross-currency basis*.

More formally, the cross-currency basis, which I denote by $\varepsilon_{\mathrm{CCB}}$, is defined as

$$\varepsilon_{\mathrm{CCB}} \equiv \frac{1}{\tau}\left(\ln f_t(T) - \ln f_t^{\mathrm{m}}(T)\right). \tag{2.3}$$

Substituting in Eq. (2.1) for $f_t(T)$, we can express this in the following, perhaps more intuitive, form,

$$f_t^{\mathrm{m}}(T) = e^{(r_d-(r_f+\varepsilon_{\mathrm{CCB}}))\tau}. \tag{2.4}$$

In words, the strike of the zero cost forward does not match that predicted using the interest rate differential alone. Instead, one must add a cross-currency basis adjustment of $\varepsilon_{\mathrm{CCB}}$.

To date, $\varepsilon_{\mathrm{CCB}}$ is typically negative. For instance, consider the example of EUR-USD again. Figure 2.2 shows that with $\varepsilon_{\mathrm{CCB}} = -0.2\%$, $r_d = 2\%$ and $r_f = 0\%$ then $f_t^{\mathrm{m}}(T) > f_t(T)$. This reflects a demand for USD funding in that market participants are prepared to buy EUR and sell USD in the future at a more unfavourable rate than that implied by the interest rate differential, perhaps in order to sell EUR and obtain USD at the present time.

The cross-currency basis has been persistently in favour of the USD since the GFC, reflecting persistent USD funding shortages. I refer interested readers to Borio et al. (2018).

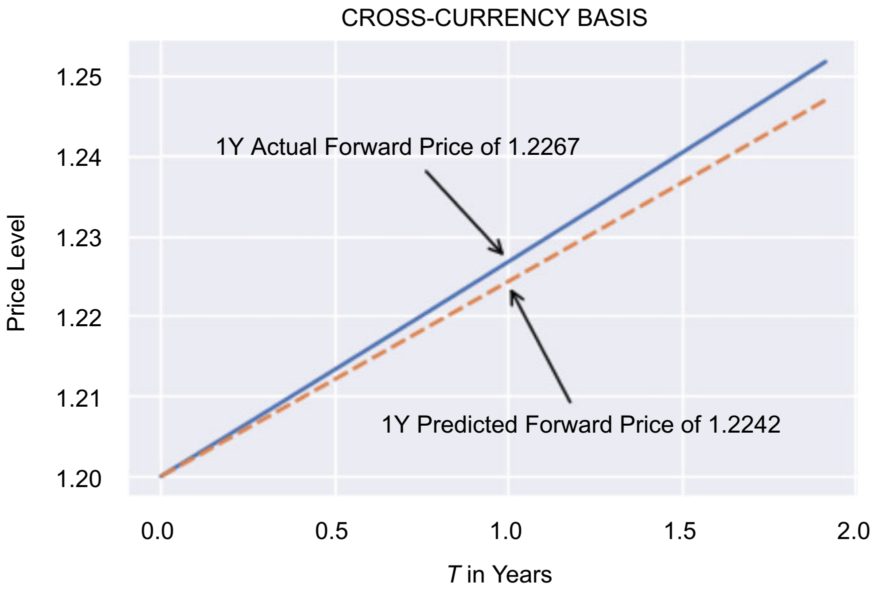


Fig. 2.2 The dashed line shows the forward based on the CIP relationship. The solid line shows the actual forward observed in the market. Note that the actual forward lays above the predicted forward. The difference between the two is called the cross-currency basis. In this example, EUR bought forward trade at a premium relative to their predicted price. This reflects a demand for USD funding now

2.2 Uncovered Interest Rate Parity (UIP) and the Carry Trade

2.2.1 Forward as the Expectation of the Future Spot Rate

Applying Eq. (1.11) to the spot FX rate gives us

$$S_t = e^{-r_d \tau} \mathbb{E}_t^P \left[e^{r_f \tau} S_T \right] + \mathrm{cov}_t \left(\beta \frac{u'(c_T)}{u'(c_t)}, e^{r_f \tau} S_T \right). \tag{2.5}$$

Here, the time period of interest is the maturity of the forward, $t + 1 = T$, and the payoff, $X_T = e^{r_f \tau} S_T$, is the final value of the spot rate plus the accumulated interest while the investment in the foreign currency sits invested in the foreign safe asset (foreign bond or bank account) for the period between time t and T.

Finally, substitute Eq. (2.5) into our expression for the forward rate as a function of the spot rate, Eq. (2.1), to find that,

$$
\begin{aligned}
f_t(T) &= \mathbb{E}_t^P[S_T] + \mathrm{cov}_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, e^{r_f \tau} S_T\right) \\
&= \mathbb{E}_t^P[S_T] + \underbrace{e^{r_f \tau} \rho_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, S_T\right) \sigma_t\left(\beta \frac{u'(c_T)}{u'(c_t)}\right) \sigma_t(S_T)}_{\text{FX Risk Premium}}. \qquad (2.6)
\end{aligned}
$$

This equation is key to understanding FX forwards. It is common among practitioners to suggest that the forward rate, $f_t(T)$, reflects the market's expectation of the future spot rate at time T. However, Eq. (2.6) makes clear that this statement is not generally true. It is only true if the FX risk premium term is zero.

The risk premium is zero if investors are risk neutral because $u'(c) = \alpha$ in this case. As discussed in Chapter 1, this is generally not the case and investors exhibit risk averse behaviour.

The risk premium is zero if the spot rate is uncorrelated with consumption. As discussed in Chapter 1, currencies exhibit strong correlations to the wider economy, and so again, this correlation is generally far from zero.

Finally, the risk premium is zero if $\sigma_t(S_T) = 0$. However, in this case, there is no uncertainty, and the concept of *risk* is no longer valid.

The forward rate should therefore be thought of as the expected future spot rate, plus a risk premium. I illustrate this in Fig. 2.3. It is this risk premium that drives the so-called carry trade. This is the topic of the next section.

2.3 The Carry Trade

The carry trade typically attempts to profit by borrowing in low interest rate currencies and investing in high interest rate currencies. This is equivalent to trading a forward contract, and then unwinding the delivered spot in the FX market upon delivery. To see this, consider the following example of a U.S.-based investor investing in Mexico.

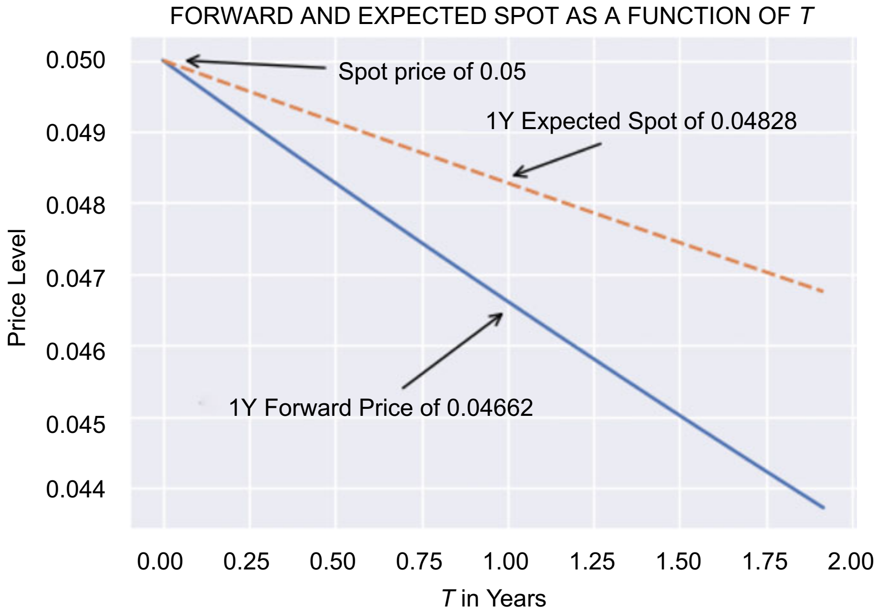


Fig. 2.3 The figure takes MXN-USD as an example with $S_t = 0.05$. The USD interest rate is $r_d = 1\%$ and the MXN interest rate is $r_f = 8\%$. Since $r_f > r_d$ we have a downward sloping $f_t(T)$ curve, shown using a solid line. The dashed line shows the expected spot path. The risk premium is the difference between the two. At the 1Y point, the forward price is 0.04662, and the expected spot value is 0.04828, providing a risk premium of 0.00166 cents, or 3.5% per annum. This is the risk premium earned by the investor by implementing a long MXN carry strategy, as discussed in Sect. 2.3 (*Note* Market practitioners may be familiar with upward sloping forward curves in USD-MXN. I remind the reader that the quotation convention in this chapter is MXN-USD, rather than USD-MXN, and this generates a downward sloping forward curve)

The MXN-USD spot rate today is S_t.[2] The investor borrows S_t USD to purchase 1 MXN and invests this into a time T maturity Mexican government bond. At time T, the investor's holding in Mexican government bonds has accumulated interest and is now worth $e^{r_f \tau}$ MXN. She sells this holding and converts her MXN back to USD at the prevailing market spot rate of S_T, to receive $e^{r_f \tau} S_T$ USD. Finally, she repays the S_t USD that she had borrowed, which has accumulated in value to $S_t e^{r_d \tau}$ after interest. The carry

[2] Market convention is to quote USD-MXN, which represents the number of MXN per USD. However, for simplicity, I maintain the notation of FOR-DOM in this chapter, with the USD as the domestic currency. Therefore, I proceed by quoting the number of USD per MXN Peso, and denote this as MXN-USD. This is simply the inverse of the USD-MXN spot rate.

trade's payoff is therefore

$$S_T e^{r_f \tau} - S_t e^{r_d \tau}. \tag{2.7}$$

Now, suppose that instead of the above, the investor simply bought $e^{r_f \tau}$ units of the forward contract on MXN-USD at time t. Her payoff is then

$$e^{r_f \tau}(S_T - f_t(T)). \tag{2.8}$$

Substituting in Eq. (2.1) for $f_t(T)$, the payoff in Eq. (2.8) becomes $S_T e^{r_f \tau} - S_t e^{r_d \tau}$, which is identical to Eq. (2.7).

The key point to note is that buying the zero cost forward contract at time t and then selling the currency that is delivered at time T at the prevailing market rate, S_T, is identical to implementing a carry strategy. An investor may implement the carry strategy if her expectation of the future spot rate is sufficiently higher than the forward rate, as is the situation in Fig. 2.3.

Expressing the carry strategy in terms of forwards as we have done here allows us to use the risk premium equations that we have already derived to better understand the relationship between UIP and the carry trade. This is the topic of the next subsection.

2.3.1 UIP, Risk Premiums and the Carry Trade

Suppose that the investor implemented the carry strategy with notional $e^{-r_f \tau}$ MXN rather than 1 MXN. The previous subsection shows that her payoff is $S_T - f_t(T)$. UIP is said to hold if

$$\mathbb{E}_t^P[S_T] - f_t(T) = 0. \tag{2.9}$$

In words, UIP is said to hold if there are no expected profits to implementing the carry strategy. On average, the spot rate moves to the forward, and therefore carry strategies yield zero profit. However, the theory that we have developed on risk premiums suggests that UIP should not hold.

To see this, note that Eq. (2.6) already provided the expected profits from carry trading,

$$\mathbb{E}_t^P[S_T] - f_t(T) = -e^{r_d T} \rho_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, S_T\right)\sigma_t\left(\beta \frac{u'(c_T)}{u'(c_t)}\right)\sigma_t(S_T). \tag{2.10}$$

Carry is therefore simply a risk premium like any other. As usual, the risk premium depends on the correlation of the spot rate with the wider economy/marginal utility. Investment in currencies that are most correlated with the state of the economy, such as those of emerging market countries, requires a large and positive risk premium. However, currencies that behave as portfolio/consumption hedges may command a negative risk premium.

To build further intuition around Eq. (2.10), let us plug MXN-USD in as an example. Recall Fig. 1.1 showed us that MXN-USD weakened during the economic turmoil associated with the Coronavirus in 2020. This was the case also during the financial crisis, and other economic crises. That is, for MXN-USD, $\rho_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, S_T\right)$ is negative because when consumption falls, marginal utility rises, and MXN-USD falls. This means that $\mathbb{E}_t^P[S_T] - f_t(T) > 0$ and therefore, in equilibrium, the market must price a positive expected return for a carry strategy that is long MXN and funded in USD. To complete the intuition around Eq. (2.10), I provide the feature box *A Practical Experiment with Carry Risk Premium Equation* that studies the equation in the context of the movements in the MXN-USD exchange rate over Q1 of 2020.

To summarise, our model of risk premiums predicts a UIP violation in favour of the USD, CHF, JPY and sometimes EUR. That is, a position that is long one of these four currencies vs short AUD, MXN (or other currencies positively correlated with consumption such as most emerging market currencies) carries a negative expected return. In the case of the USD, this has been famously referred to as its *exorbitant privilege*. I discuss this idea in detail in the feature box titled *Exorbitant Privilege and Risk Premiums* later in the chapter.

2.3.1.1 Practical Consideration on Implementation of the Carry Trade

An important practical consideration for investors implementing carry trades is to note that Eq. (2.10) is an equilibrium equation. It applies only after investors have finished adjusting their portfolios to satisfy their first-order condition (Eq. 1.6). Therefore, during bouts of risk aversion, it is important to enter the carry trade only after sufficient repricing in the spot rate has occurred to account for the higher values of $\rho_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, S_T\right)$, $\sigma_t\left(\beta \frac{u'(c_T)}{u'(c_t)}\right)$ and $\sigma_t(S_T)$ that are typically associated with market sell-offs. For instance, Fig. 2.4 shows that it took a period of several trading sessions for $\sigma_t(S_T)$ in

Fig. 2.4 The figure shows how the spot price of MXN-USD, and the market implied volatility extracted from FX options evolved following the outbreak of the 2020 coronavirus pandemic. The news of the pandemic hit the market in early March. However, it was not until 18 March that MXN-USD implied volatility hit its peak and MXN-USD spot hit its low. This suggests that it can take some time for markets to reprice to the point where the equilibrium condition of Eq. (2.10) is satisfied (*Source* Data from Bloomberg LLP, with permission)

MXN-USD to sufficiently reprice during the Coronavirus crisis. The crisis began to hit markets at the beginning of March. However, it was not until March 18, 2020 that the full repricing in the spot rate and implied volatility was observed. The precise timing at which Eq. (2.10) holds is therefore not clear. However, it is clear that the most immediate mistake an investor can make is to assume that markets adjust instantaneously or quickly.

A Practical Experiment with the Carry Risk Premium Equation

I carry out a back-of-the-envelope implementation of Eq. (2.10) to test its approximate performance over the Coronavirus crisis in 2020. To begin, I divide through by S_t and express Eq. (2.10) in gross returns as

$$\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right] - \frac{f_t(T)}{S_t} = -e^{r_d\tau}\rho_t\left(\beta\frac{u'(c_T)}{u'(c_t)}, \frac{S_T}{S_t}\right)\sigma_t\left(\beta\frac{u'(c_T)}{u'(c_t)}\right)\sigma_t\left(\frac{S_T}{S_t}\right).$$

Next, I study how the terms in this equation change across two points in time, 2 Jan, before the pandemic hit financial markets, and 18 March, at the depth of the crisis. I label 18 March as the depth of the crisis because this is the date of the highest reading of 1Y implied volatility of options on the S&P500 index.

First, I estimate the change in $\sigma_t\left(\frac{S_T}{S_t}\right)$ using market implied volatility in USD-MXN from 1Y expiry FX options. Implied volatility increased from 8.6% on 2 Jan 2020, to 18.2% on 18 Jan 2020, a factor of 2.2.[3]

Estimating $\sigma_t\left(\beta\frac{u'(c_T)}{u'(c_t)}\right)$ is more difficult. To do so, I apply Eq. (1.37) that we used to derive the CAPM, and take the S&P 500 as a proxy for R^M. Taking the volatility on both sides gives us $\sigma_t\left(\beta\frac{u'(c_T)}{u'(c_t)}\right) = b\sigma_t(R_T^M)$. S&P 500 implied volatility increased from 15.2% on 2 Jan to 46% on 18 March, and therefore $\sigma_t\left(\beta\frac{u'(c_T)}{u'(c_t)}\right) = b\sigma_t(R_T^M)$ increased by a factor of 3.0.

Next, r_d fell by approximately 1% as the Federal Reserve instituted emergency interest rate cuts in response to the turmoil. Therefore, $e^{-r_d\tau}$ fell by a factor of 1.01. This is small enough to ignore.

3 Readers unfamiliar with the concept of implied volatility may refer to Chapter 1 of Iqbal (2018).

Finally, although $\rho_t\left(\beta \frac{u'(c_T)}{u'(c_t)}, \frac{S_T}{S_t}\right)$ likely became larger in magnitude (more negative), I continue the calculation assuming that it was unchanged. This provides a conservative estimate of the change in risk premium. The table below summarises the changes.

Term	2 Jan 2020	18 March 2020	Factor Change
$\sigma\left(\frac{S_T}{S_t}\right)$	8.6%	18.2%	2.2
$\sigma\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}\right)$	$b \times$ 15.2%	$b \times$ 46%	3
$\rho\left(\beta \frac{u'(c_{t+1})}{u'(c_t)}, \frac{S_T}{S_t}\right)$	x	x	1

Taking the product of the factor changes in the table, I find that the risk premium/expected return in MXN-USD increased by a factor of approximately 6.6. If, for example, the value of $\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right] - \frac{f_t(T)}{S_t}$ calculated with $T = 1$ year and at $t = 2$ Jan was 4%, our estimate of this quantity calculated at $t = 18$ Jan is $6.6 \times 4\% = 26.4\%$.

For the expected return of a long MXN-USD position to increase to 26.4%, either the forward price of MXN-USD must cheapen in that $\frac{f_t(T)}{S_t}$ must be much lower on 18 March than on 2 Jan, or, the expected spot return, $\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right]$ must increase, or both.

A Practical Implementation of the Carry Risk Premium Equation

Studying the data, it is clear that the forward price of MXN did not cheapen. $\frac{f_t(T)}{S_t}$ remained at around 0.93. Mexican interest rates did not increase to provide the required risk premium, and remained approximately 7% higher than U.S. interest rates throughout the crisis period. Therefore, $\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right]$ must have increased.

To generate the 26.4% expected return, $\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right] - 1$ needed to reprice to 19.4% (with the remaining 7% coming from the forward, as discussed). It is unlikely that investors would price $\mathbb{E}_t^P[S_T]$ higher, and therefore the change must come from a fall in S_t. If $\mathbb{E}_t^P[S_T]$ remained unchanged, then S_t must fall by 19.4%. If $\mathbb{E}_t^P[S_T]$ fell by, say, 5% then S_t was required to fall by 24.4%.

The MXN-USD rate on March 18, 2020 was in fact 25% lower than its Jan 2, 2020 value. This ties out loosely with the quantities studied in this feature box and suggests that the model does a reasonable job at explaining the data.

The reader will note that the calculations above are predicated on an expected return of a MXN-USD carry trade of 4% on 2 Jan. Here, I argue that this is reasonable. Since inflation in Mexico is typically of the order of 3% higher than in the U.S., and the market value of $f_t(T)/S_t$ was 0.93 (7% discount to spot), it is reasonable to suggest that $\mathbb{E}_t^P\left[\frac{S_T}{S_t}\right] - 1$ was of the order of −3%, to leave an expected return of 4% for a carry trader. I provide more detailed justification for this argument in Chapter 3, where we study the relationship between inflation and real exchange rates.

Finally, there are several known sources of bias in the above calculation. First, $\rho_t\left(\beta\frac{u'(c_T)}{u'(c_t)}, \frac{S_T}{S_t}\right)$ likely became more negative. This suggests that the fall in MXN-USD should have been even greater than 25%. However, offsetting this one may argue that using 1Y USDMXN and S&P 500 implied volatility overestimate the changes in $\sigma_t\left(\frac{S_T}{S_t}\right)$ and $\sigma_t\left(\beta\frac{u'(c_T)}{u'(c_t)}\right)$ respectively because these values contain an increasing volatility risk premium. Nevertheless, the calculations here provide a useful starting point for FX investors to understand the effect of changing risk premiums on prices.

2.3.2 Carry, Carry Trading and Interest Rate Differentials

Readers familiar with FX carry trading may note that the strategy's expected profits according to the right-hand side of Eq. (2.10) do not directly depend on the interest rate differential, $r_d - r_f$. Since interest rate differentials are often one of the main considerations for investors involved in such strategies, this may be surprising. To understand why, it is important to carefully distinguish between what investors typically refer to as *carry* and expected profits in our equilibrium model of an investor portfolio.

Carry refers to the hypothetical profit that would be earned from holding a currency position over a period of time, when spot rate ends that time period unchanged. It may change during the time period, but *carry* refers to the case where spot returns to the point at which it began.

Recall from the previous section that if the investor implements the carry strategy with notional $e^{r_f \tau}$ then her payoff is $S_T - f_t(T)$. If the spot rate remains unchanged between time t and T, then $S_T = S_t$ and the investor's profit is

$$\begin{aligned} S_t - f_t(T) &= S_t - S_t e^{(r_d - r_f)\tau} \\ &= S_t\left(1 - e^{(r_d - r_f)\tau}\right) \\ &\approx S_t(r_f - r_d)\tau. \end{aligned} \tag{2.11}$$

Her return (profit per unit of domestic currency invested) is therefore approximately $(r_f - r_d)\tau$, which is the formula that carry investors are familiar with. For example, if interest rates on the MXN and USD are 8%, and 1% respectively, then the carry earned on a long position in MXN funded in USD is approximately 7% per USD. The key difference between Eq. (2.10) and the carry formula, Eq. (2.11), is that the former comes from an equilibrium model of investor behaviour—it does not assume that the spot rate will remain fixed.

As long as the spot rate depreciates at a rate slower than the carry, the investor implementing the carry trade makes money. In our example, an investor into MXN requires MXN to depreciate against the USD by anything less than 7% per year to make money. Therefore, investors may consider the annual carry calculated using Eq. (2.11) as their *buffer* against adverse movements in the spot rate over the course of the investment. Carry is therefore not a measure of expected profits, but remains a useful heuristic in carry trading.

Finally, it is interesting to consider why the market interest rate for MXN is higher than that for USD. There are several possible reasons. For example, higher interest rates may reflect sovereign credit risks, restrictions on capital flows, risk premiums in the interest rate term structures, among others. However, one of the most important reasons for the purposes of FX investors is relative inflation differentials. I study this idea in Chapter 3.

Let us now recapitulate the key points. The expected profit from carry trading is the difference between the expected future spot rate and the forward, namely $\mathbb{E}_t^P[S_T] - f_t(T)$. $f_t(T)$ is determined by the interest rate differential between the domestic and foreign currencies. The investor may calculate $\mathbb{E}_t^P[S_T]$ using their preferred process. However, this book argues that it should contain a risk premium. The risk premium model developed in this book suggests that the expected profit to carry trading is given by Eq. (2.10).

Exorbitant Privilege and Risk Premiums

In the 1960s, French finance minister (eventually president) Valéry Giscard d'Estaing famously bemoaned the U.S. Dollar's "exorbitant privilege". Eichengreen (2011) interprets this commonly used phrase:

> With cheap foreign financing keeping U.S. interest rates low and enabling American households to live beyond their means, poor households in the developing world ended up subsidising rich ones in the U.S. The cheap finance that other countries provided the U.S. in order to obtain the dollars needed to back an expanding volume of international transactions underwrote the practices that culminated in the crisis. The U.S. lit the fire, but foreigners were forced by the perverse structure of the system to provide the fuel.

In the language of risk premiums, "exorbitant privilege" means that there is a UIP violation that favours the USD. This can manifest itself through U.S. interest rates being low compared with other nations, and/or the price of the USD being high. Holding long USD provides negative returns on average.

Equation (2.10) tells us that the UIP violation in favour of the USD is because the USD is negatively correlated with the wider economy and acts as a portfolio hedge. In Chapter 1, I provided a sample period, namely the Coronavirus pandemic of 2020, to illustrate this negative correlation in the USD, JPY and CHF. However, I did not discuss the reasons that such correlations exist. This is the topic of the remainder of this feature box.

1. Safe Asset Demand

Demand for savings and safe assets increases during recessions and periods of increased economic uncertainty. I showed this through a formal model of risk averse investors in Sect. 1.6. This may lead to demand for USD, JPY, CHF and EUR because these denominations are able to provide such safe assets. The case in favour of the USD is particularly strong because U.S. treasury markets are the largest and most liquid in the world, and denominated in the world's foremost reserve currency. The U.S. is able to satisfy the world's demand in such circumstances to a greater extent than its competitor currencies and this may lead to a strong bid for USD. Prasad (2015), for instance, provides extensive analysis on this and related arguments.

The component of this argument that is perhaps unsatisfying is that it suggests that the USD should be *more* negatively correlated with the economy than CHF and JPY, because of the greater depth of availability of U.S. safe assets. However, it is well known that USD-CHF and USD-JPY often weaken during crisis periods. I suggest some further reasons for this later in the chapter.

2. USD Invoicing in Global Trade

In the quotation above, Eichengreen (2011) alludes to another potential reason for the UIP violation in favour of the USD—the fact that outsized quantities of international trade is invoiced in USD relative to U.S. exports.

Gopinath and Stein (2018) provide a mechanism through which USD invoicing may lead to a UIP violation. The authors identify a loop. If trade is invoiced in USD, then non-U.S. importers demand safe USD-denominated assets to hedge their consumption risk. The U.S. Treasury can supply such safe assets, and U.S. banks may be able to manufacture them. However, if the demand remains bigger than the supply from U.S. institutions, local banks need to manufacture these assets, but at a natural comparative disadvantage. They will therefore only do so at a low interest rate. This then increases the incentive to invoice in USD because such invoicing generates the collateral needed to borrow cheaply in dollars, and the loop continues.

3. Carry Trade Unwind

Carry trade strategies are often unwound during crises as investors de-risk and this is often cited as a reason for the strength of EUR, JPY, CHF and USD during times of economic turbulence. However, this argument is somewhat circular because it may already be the case that it is precisely *because* EUR, JPY, CHF and USD are portfolio hedges, that they pay lower interest rates in the first place and thereby become candidates for the funding leg of a carry trade.

Furthermore, it should be noted that AUD and USD interest rate differentials had already fallen to zero preceding the economic turbulence associated with the Coronavirus pandemic in February of 2020. However, despite this, the AUD fell significantly against the USD. The carry trade argument is not able to explain this observation.

4. Net International Investment Positions (NIIPs) and Current Account Balances
Perhaps a country's NIIP or current account balance determines its currency's behaviour during economic crises? The NIIP idea is that repatriation flows and/or currency hedging take place that strengthen the currencies of countries with positive NIIP and weaken the currencies of countries with negative NIIP. Phrased in terms of current accounts, the idea is that a country running a current account deficit may struggle to finance this during economic turmoil, and the currency must weaken against current account surplus countries to redress the imbalance. In practice, this idea has not held up well, particularly in the case of the USD. For example, prior to the global financial crisis (GFC), the U.S. ran a large current account deficits of over 5% of U.S. GDP. However, contrary to the predictions of many notable economists at the time, the subsequent financial crisis that led to USD strength rather than weakness, despite the fact that it emanated from the U.S. mortgage market. Perhaps this is the reason why the term "exorbitant" is often used to describe the USD's violation of UIP, whereas this term is not used to describe the even larger CHF and JPY violations, because Switzerland and Japan have not run deficits to the size of the U.S. and these currencies are therefore, in some sense, more deserving of their respective UIP violations.

I provide an in-depth discussion of current and financial accounts, the NIIP and other macroeconomic accounts that may influence the FX rate in Chapter 5.

2.3.3 Analogy with Equity Investments

It is noteworthy that all of the equations derived in this chapter apply equally to equity investments. Simply replace r_f with the dividend rate on an equity investment, and Eq. (2.10) applies. The term *carry trade* is typically used to describe an FX strategy, but as far as financial economic theory is concerned, it is no different to, for example, borrowing in USD to fund an investment in the S&P500. Just as in FX, such an investment is equivalent to buying forward (for futures) contracts on the S&P500.

The S&P500 is quoted in USD. In FX nomenclature, we may write its price as S&P500-USD, and think of S&P500 as the foreign currency and the USD as the domestic currency, and then apply the equations in this chapter as we have in the context of FX. It is now clear that what market participants refer to as FX carry is the same in economic terms as the well known equity risk premium. Both are risk premiums that are earned by investors for exposing themselves to risks that are correlated with their consumption streams.

2.4 Chapter Summary

- $f_t(T)$ is determined by the interest rate differential between the domestic and foreign currencies. This is known as CIP. However, since the GFC, CIP does not hold perfectly. The small but persistent deviation from CIP is known as the cross-currency basis.
- The expected profits from carry trading are the difference between the expected future spot rate and the forward, namely $\mathbb{E}_t^P[S_T] - f_t(T)$. This is the carry risk premium.
- The investor may calculate $\mathbb{E}_t^P[S_T]$ using their preferred process. However, this book argues that it should contain a risk premium. The risk premium model developed in this book suggests that expected profits to carry trading are given by Eq. (2.10).
- UIP is the theory that the risk premium in currencies is zero. This does not hold in practice. The USD and other so-called safe-haven currencies carry a negative risk premium. Meanwhile, emerging market and some developed market currencies carry a positive risk premium.

References

Borio, C., Iqbal, M., McCauley, R., McGuire, P., & Sushko, V. (2018). *The failure of covered interest parity: FX hedging demand and costly balance sheets* (BIS Working Papers 590).

Eichengreen, B. (2011) , *Exorbitant privilege: The rise and fall of the dollar*. OUP Oxford.

Gopinath, G., & Stein, J. C. (2018). *Banking, trade and the making of a dominant currency* (NBER Working Paper 24485).

Iqbal, A. (2018). *Volatility*. Wiley Finance Series: Practical Options Theory.

Prasad, E. (2015) , *The dollar trap: How the U.S. Dollar tightened its grip on global finance*. Princeton University Press.

3

Exchange Rates, Interest Rates, Inflation and the Risk Premium

Exchange rates respond to changes in interest rates, inflation expectations and the risk premium. This chapter introduces fixed real exchange rate models, of which the well known Purchasing Power Parity (PPP) model is a special case that sets the real exchange rate to 1.0. Such models relate the expected future exchange rate to relative inflation expectations. Combining this idea with the concept of CIP from Chapter 2, the reader is able to understand the relationship between the FX spot rate, nominal interest rates and inflation expectations.

Before deriving a formal mathematical relationship between the aforementioned variables, the chapter builds intuition by proceeding through five practical examples.

Example (i) shows that higher inflation expectations lead to movement in the forward price of a currency. However, if the nominal interest rate is able to price in the higher inflation, then the real interest rate and spot FX price need not change.

Example (ii) discusses the case where higher inflation is not met with a higher nominal interest rate. Such a situation may occur due to policies pursued by the central bank, such as yield curve targeting, or quantitative easing (QE). In this circumstance, I show that higher inflation is met with a fall in the real interest rate, and a weakening in both the spot and forward FX price.

Example (iii) considers an unexpected hike in the interest rate that is instigated by a central bank with a view of cooling the economy. I show that

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_3

this circumstance can lead to so-called *exchange rate overshooting* as the real interest rate rises.

Example (iv) studies concurrent movements in the central bank policy rate, inflation expectations and interest rates, as often occur in practice and shows the reader how to aggregate such effects to calculate FX spot and forward price movements.

Finally, in Example (v) I refocus on the risk premium, which we studied in general in Chapter 1 and in the context of UIP in Chapter 2. I combine this with the fixed expected real exchange rate model and CIP relationship to determine how currencies move when the risk premium is itself linked to inflation expectations, a situation that is pertinent in emerging market currencies.

Let us begin by understanding the meaning of the real exchange rate, PPP and the real interest rate.

3.1 Fixed Real Exchange Rate Models

3.1.1 Defining the Real Exchange Rate

With the nominal spot rate, S_t, defined as FOR-DOM (i.e. number of units of domestic currency, per unit of foreign currency), the real exchange rate, e_t, is defined as,

$$e_t = S_t \times \frac{P_{t,f}}{P_{t,d}}, \tag{3.1}$$

where $P_{t,f}$ and $P_{t,d}$ denote the price levels in the countries of the FOR and DOM currencies respectively.

To avoid confusion, let us consider an example involving the real bilateral exchange rate of EUR versus USD. For simplicity, I assume that there is a single type of real consumption good, called widgets. Suppose that a widget costs 150 EUR in Europe and 100 USD in the U.S. and that the nominal EUR-USD exchange rate is 1.20. Since the quotation convention here is EUR-USD, EUR is FOR and USD is DOM. The real exchange rate is therefore

$$\begin{aligned} e_t &= S_t \times \frac{\text{Price of widgets in Europe in EUR}}{\text{Price of widgets in United States in USD}}, \\ &= 1.2 \text{ USD per EUR} \times \frac{150 \text{ EUR}}{100 \text{ USD}} = 1.8. \end{aligned} \tag{3.2}$$

This equation tells us that widgets are expensive in Europe compared to the U.S. To see this, note that the owner of a European widget can sell this widget for 150 EUR, then sell those 150 EUR in exchange for 180 USD, and finally use the 180 USD to buy 1.8 American widgets.

More generally, the real exchange rate compares the prices of a broad basket of goods in different countries. A real exchange rate of 1.8 means that average consumer prices in Europe at 80% higher than in the U.S.

3.1.2 PPP and the Law of One Price

Consider a hypothetical global economy in which each country produces identical widgets only and they are frictionlessly transportable across borders to be traded. The *law of one price* tells us that widgets must trade at the same price globally. Applied in the context of international trade, this is called PPP. The implication is that the real exchange rate cannot be anything other than 1.0. To see how frictionless trade leads to a real exchange rate of 1.0, consider the following.

If the real exchange rate were 1.8, as in the example in the previous subsection, then an arbitrageur may exploit the following trading strategy. First, borrow 100 USD and buy a widget in the U.S. Then, frictionlessly transport the widget to Europe, and sell it to receive 150 EUR. Finally, sell the 150 EUR in exchange for $150 \times 1.2 = 180$ USD, repay the 100 USD originally borrowed and make a risk free profit of 80 USD. Arbitrageurs will continue to exploit this strategy until the real exchange rate reaches 1.0. This can happen through any of, or a combination of, the following three channels. First, the demand from arbitrageurs for U.S. widgets can drive their price higher. Second, the supply from arbitrageurs of widgets in Europe can drive their price lower. Third, the selling of EUR-USD in the FX market can drive the nominal exchange rate lower. The next feature box discusses PPP in practice.

PPP in Practice

The Economist, a newspaper, has famously published a measure of the real exchange rate known as the Big Mac Index since 1986.[1] As its name suggests, the widget is McDonald's Big Mac burger. The project shows considerable deviations from PPP over time. For example, in 2000, the Big Mac cost approximately 20% more in the United Kingdom (U.K.) than in the U.S., but by 2020, it cost approximately 20% less. That is, the real exchange rate fell from approximately 1.2 to 0.8 over the past 20 years.

More generally, economists and policymakers are more interested in a related concept, the real effective exchange rate (REER). The REER is an average of the bilateral real exchange rates between the country and each of its trading partners, weighted by the respective trade shares of each partner. Since it is an average, a country's REER may display no overall misalignment when its currency is overvalued relative to that of one or more trading partners so long as it is undervalued relative to others.

Figure 3.2 shows how REERs have changed since 2010. It is clear that there is considerably volatility in the REER. For instance, the REERs of Australia, Canada, Japan, Norway and South Africa have fallen by approximately 20%, while those of China and the U.S. have risen by a similar amount. The REER of the Switzerland and the U.K. is largely unchanged (Fig. 3.1).

The PPP approach suggests that real exchange rates and therefore REERs should not exhibit volatility and therefore these data reject the PPP approach to some degree. Nevertheless, understanding how REERs have changed over time remains a useful input for investors into currency valuation.

The feature box showed that real exchange rates do deviate away from 1.0 in practice. Why might this happen? The most obvious reason is transport friction. For PPP to hold, arbitrageurs are required to transport misaligned widgets across borders, as described in Sect. 3.1.2. Since transport is costly, and restrictions to trade such as taxes and tariffs exist, it is not possible to execute this arbitrage across all goods.

A second reason may relate to data measurement. Even if all goods were frictionlessly transportable, the consumer price index (CPI) baskets that are used to calculate the REER in Fig. 3.2 are not identical to each other (I

[1] The index is available at https://www.economist.com/big-mac-index.

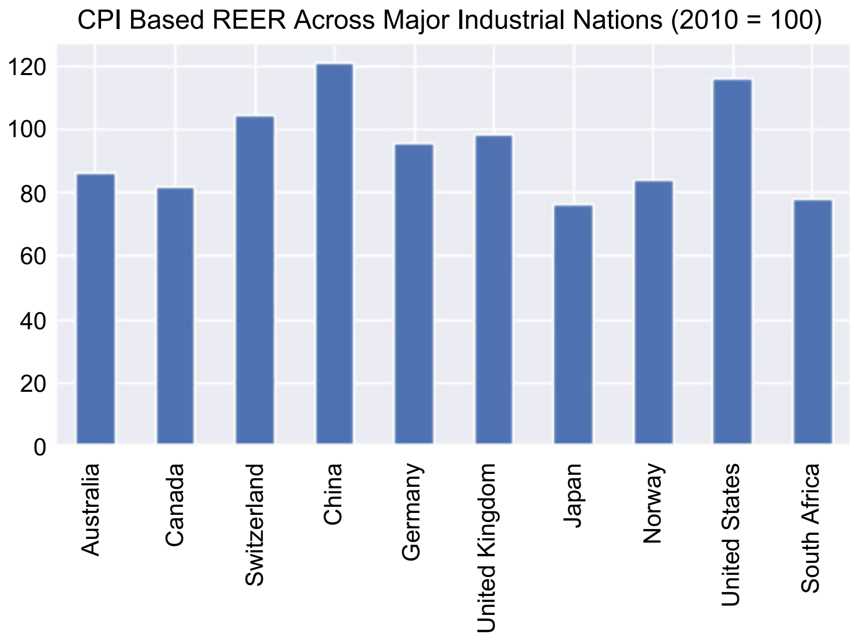


Fig. 3.1 The REER is an average of the bilateral RERs between the country and each of its trading partners, weighted by the respective trade shares of each partner. The figure shows how the REER has changed across 10 industrialised countries since 2010. The calculation normalises the price of the basket of goods that underlay the CPI index to 100 in 2010. The plot shows the price of the basket of goods in 2019 (*Source* World Bank Data Catalogue)

provide a detailed discussion of the CPI basket in Chapter 7). There is no reason to suppose that changes in the FX rate should offset differences in the officially chosen cross section of measures of inflation across nations.

Third, internationally traded goods may not be of the same quality. A European car may be of higher or lower quality on average than an American car, and this may lead to the REER deviating in favour of the country that produces the higher quality good. This idea is closely related to a change in the *terms of trade*, which I discuss in Sect. 3.2.

Fourth, not all goods are tradeable. For example, cars, computers, some services and other widgets are easily transportable across borders, but haircuts, gardening services, and many other (often lower tech services) are not. The next section discusses how this fact may impact PPP.

Fig. 3.2 The figure shows a steady increase in productivity in China since 1995, proxied by GDP per capita measured in 2010 USD (left axis). The figure also shows a trend higher in the value of Chinese Yuan's REER (right axis). The data provides some support for the Balassa-Samuelson theory, which predicts a strengthening in the REER as productivity increases (*Sources* Bank for International Settlements, Real Broad Effective Exchange Rate for China [RBCNBIS]. and World Bank, Constant GDP per capita for China [NYGDPPCAPKDCHN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/RBCNBIS, January 30, 2021)

3.1.3 The Balassa-Samuelson Theory

In addition to a deviation in the real exchange rate away from 1.0, a further empirical finding is that the real exchange rate is lower in poor countries rather than rich ones. That is, a single USD buys more widgets in a poorer nation than it does in a rich one. Balassa (1964) and Samuelson (1964) attempted to explain this phenomenon.

The authors assume that productivity (output per worker hour) is higher in developing countries in the tradeable sector. For example, better technology means that workers in developed countries may produce more cars per worker hour. However, productivity differences in non-tradeables across developed and developing countries are negligible—a gardener may mow a similar number of gardens in a developed or developing country.

The higher productivity in the tradeables sector of developed countries leads to higher real wages for workers in both the tradeables and non-tradeables sector. These higher wages lead to higher production cost and higher prices in developed countries. This causes a PPP deviation in the direction of raising the real exchange rate of the developed country.

The secular horizon impact of the Balassa-Samuelson theory is that countries that develop technology at a faster rate should experience an increase in the real exchange rate of their currency. Figure 3.2 provides some empirical support for the Balassa-Samuelson theory by studying the REER of the Chinese Yuan (CNH) and comparing it to a measure of Chinese productivity.

3.2 Terms of Trade

Terms of trade is the ratio between a country's export prices and its import prices. Market participants are familiar with the terms of trade impacting the FX rate; if the prices of a country's exports rise relative to the prices of its imports, then that country's currency typically strengthens. How can we understand this relationship in the context of the widget/PPP economy that we have discussed in the chapter so far?

Let us again take the example of Europe and the U.S. Suppose that the initial situation is that the real exchange rate is 1.0 because arbitrageurs have traded to the point that PPP holds, as described in Sect. 3.1.2. For example, suppose that a European widget costs 100 EUR, a U.S. widget costs 120 USD, and the nominal EUR-USD FX rate is 1.20.

Next, suppose that for some exogenous reason, the world's consumers develop a preference for European widgets to the point that they are willing to substitute two U.S. widgets for one European widget. How will this impact the FX rate? An international arbitrageur should borrow 120 USD, and sell in the FX market to obtain 100 EUR. She will then purchase a European widget, export it to the U.S., and trade it for 2 U.S. widgets to obtain 240 USD. After paying back the borrowed USD amount, she is left with a profit of 120 USD. She should continue until the real FX rate moves from 1.0 to 2.0 and the profit from the strategy disappears. This may happen either by the price of European widgets in EUR increasing, the price of U.S. widgets decreasing, or by the nominal EUR-USD FX rate increasing. Often, prices are sticky or, at least, slower moving than the FX rate, and therefore the FX rate rises. This is a term of trade impact.

Real economies involve multiple goods. However, the arguments here still hold. For example, the reader may think of European widgets as cars, and U.S. widgets as computers, and the consumer preference for cars rising relative to computers.

3.3 Real Interest Rates and Expected Inflation

The interest rate observed in the market is known as the nominal interest rate. To understand the relationship between exchange rates and the nominal interest rates, we must first take a more detailed look at its components. These are known as the *real* interest rate and *expected inflation*. The remainder of this section refers to the DOM country and currency. The d subscripts may be replaced by f to denote the FOR country and currency.

Expected inflation is denoted by π_d. It is the annual continuously compounded rate at which investors expect the price level to rise. In the examples that follow, this will be the annual pace of price increases of widgets. In real markets, measures such as the consumer price index (CPI) measure the rate at which the prices of a basket of consumer goods rise (see Chapter 7). More formally,

$$\mathbb{E}_t^P\left[\ln \frac{P_{d,T}}{P_{d,t}}\right] = \pi_d \tau. \tag{3.3}$$

The real interest rate is the rate of increase in the number of units of the consumption good that an investor gains by choosing to invest in a risk free bond and consume in the future, rather than consume today. It is defined through the following equation,

$$r_d^{\text{real}} \equiv r_d - \pi_d, \tag{3.4}$$

where r_d is the nominal interest rate. This equation is often re-arranged as,

$$r_d = r_d^{\text{real}} + \pi_d, \tag{3.5}$$

and referred to as the Fisher equation, after economist Irving Fisher (1867–1947).

Equation (3.4) is perhaps best understood through an example. First, suppose that $r_d = 5\%$ and $\pi_d = 3\%$. A widget today costs 100 USD. Instead of consuming a widget today, an investor may purchase the 1 year U.S. government bond. At maturity, she receives $100 \times e^{0.05\times 1} \approx 105$ USD.[2] If the inflation rate turns out to be as expected, then the widget costs $100 \times e^{0.03\times 1} \approx 103$ USD. She may therefore consume approximately 2%

[2] Government bonds are considered risk free for the purposes of this chapter.

extra widgets in 1 year relative to what she could today. Her (continuously compounded) real rate of return is thus 2%.

3.4 FX Price Dynamics

The remainder of this chapter studies the response of FX prices to changes in r, π^e and r_d^{real} through selected examples. To disseminate the intuition as simply as possible, examples 1 to 4 assume that investors are risk neutral, $u'(c) = \alpha$, for a constant α. Therefore, the risk premium is zero, and UIP holds meaning that,

$$\mathbb{E}_t^P[S_T] = f_t(T). \tag{3.6}$$

This is simply Eq. (2.10) with the risk premium set to zero. In words, the forward is the expected future spot price. I reintroduce risk premiums in case 5 and discuss how they may influence the results that I establish here.

I assume that PPP holds in expectation at the 1 year horizon. However, the movements in the exchange rate are analogous as long as the real exchange rate is stable, even if it is so at a price other than 1.0 imposed by the PPP approach. That is, the results that follow are valid for all models that in which the real exchange rate is expected to remain stable.

The 1 year point is chosen arbitrarily and to illustrate the concepts that follow. The reader may suspect that real exchange rates are stable at time horizons longer than 1 year. This empirical question is beyond the scope of this chapter.

3.4.1 Example (i): FX Response to an Inflation Surprise with Fixed Real Yields

Consider a starting situation where PPP holds and the EUR-USD spot price is 1.20. If a widget costs 100 EUR in Europe, it must cost 120 USD in the U.S. I assume an initial situation whereby nominal interest rates are $r_f = 0\%$ in Europe and $r_d = 2\%$ in the U.S., and expected inflation rates are $\pi_f = 0\%$ in Europe and $\pi_d = 2\%$ in the U.S. Therefore, by the Fisher equation, $r_d^{\text{real}} = r_f^{\text{real}} = 0\%$. The forward path is shown in the dashed line in Fig. 3.3.

Next, the U.S. experiences a positive inflation surprise that causes market participants to revise higher their expectations of U.S. inflation from $\pi_d = 2\%$ to $\pi_d = 3\%$. This could be in response to a particularly strong U.S. jobs report, for instance. I exaggerate the change inflation expectations to make

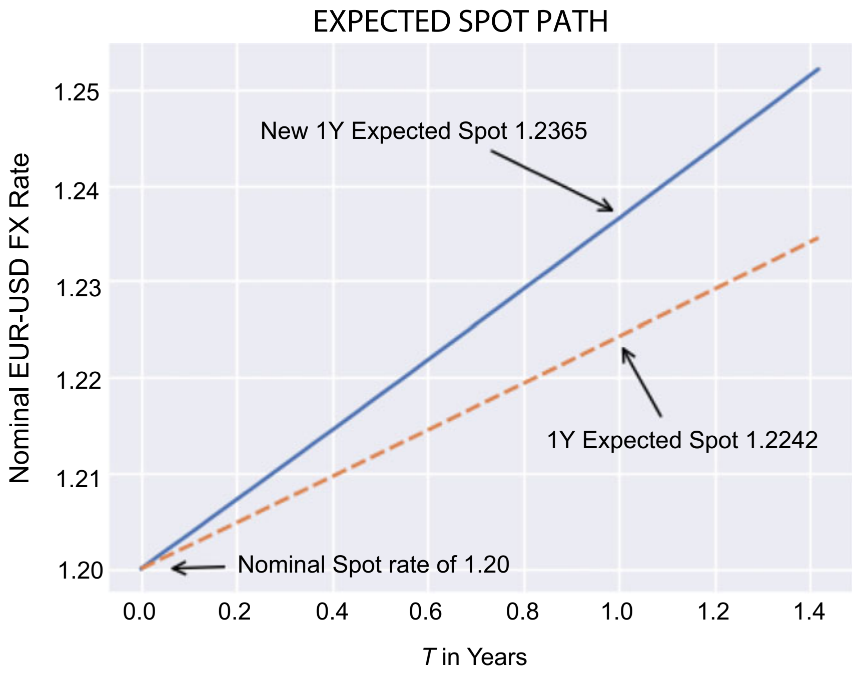


Fig. 3.3 The dashed line shows the initial situation. The nominal FX rate is 1.20. An increase in expected inflation in the U.S. means that the USD is expected to be weaker in the future, and therefore the 1 year expected forward rate must rise. This cheapening in the future value of the USD offsets the fact that prices in the U.S. will be higher in the future and keeps the expected real exchange rate fixed at 1.0. This explains why the solid line is above the dashed line. However, since the U.S. interest rate moves higher to reflect higher inflation expectations, the CIP relationship tells us that the forward path must steepen. The nominal FX rate therefore remains unchanged

the discussion clear. In typical market circumstances, a jobs report may move inflation expectations by the order of several basis points rather than 1%. How should the FX market respond to the surprise?

First, if PPP is to hold at the 1 year horizon, then the forward price of EUR-USD must be higher. With 3% inflation, U.S. widgets are expected to cost 123.65 USD in 1 year, rather than 120 USD (a 3% increase). However, since inflation in Europe will remain at 0%, European widgets will still cost 100 EUR in 1 year. For the expected real exchange rate to remain at 1.0 in 1 year, the expected nominal FX rate must move 3% higher to 1.2365. That is, USD bought 1 year forward must be cheaper relative to EUR to compensate investors for U.S. price rises in the real consumption good.

Next, since this example assumes fixed real interest rates (of zero in both the U.S. and Europe) the Fisher equation tells us that nominal interest rates move higher to fully offset the move higher in inflation expectations. Assuming that r_d moves from 2 to 3%, the CIP relationship (Eq. 2.1) tells us that the FX spot rate remains unchanged at 1.20, since it must remain 3% lower than the 1Y forward to reflect the 3% difference in nominal interest rates between the U.S. and Europe. This is shown in the solid line in Fig. 3.3.

Example (i) Summary

To summarise, the higher expected inflation in the U.S. is priced into a higher forward EUR-USD nominal FX price and into the U.S. nominal interest rates through a higher value of π_d. r_d^{real} and the spot FX rate remain unchanged.

The loose heuristic for practical trading is that if real interest rates do not change, then the spot FX price should also remain unchanged. Higher expected inflation is reflected in the price of the forward.

The reader may note that I assumed that $r_d^{\text{real}} = r_f^{\text{real}} = 0\%$. It was not necessary to assume that real interest rates are 0%. The example is easily reworked to draw the same conclusions at any level of real interest rates. However, it was necessary to assume that $r_d^{\text{real}} = r_f^{\text{real}}$. Without this assumption, it is not possible for PPP to hold at both time points. This will become clear in the next example, where I consider a case in which nominal interest rates do not move to cover higher inflation expectations.

Finally, there is no asymmetry in this example in that an inflation shock 1% lower leads to the 1 year forward price of EUR-USD falling by 1%, and the FX spot price remaining unchanged.

A Note on FX Volatility Term Structure

An inflation shock of 1% leads to the 1 year forward rising by 1%, but the spot price remaining unchanged. That is, the volatility of the forward is high relative to the volatility of spot. This dynamic suggests that implied volatility term structures should be upward sloping. A term structure of implied volatility that is not sufficiently upward sloping may therefore present a trading opportunity if the trader believes that example (i) appropriately describes the forthcoming FX price dynamics, all else being equal.

3.4.2 Example (ii): FX Response to an Inflation Surprise with Fixed Nominal Yields

The starting situation is as in example (i), and this is shown by the dashed line in Fig. 3.4. As before, an increase in U.S. inflation expectations causes the 1 year forward price of EUR-USD to move 1% higher from 1.2242 to 1.2365. However, the nominal interest rate is fixed, meaning that U.S. r_d is

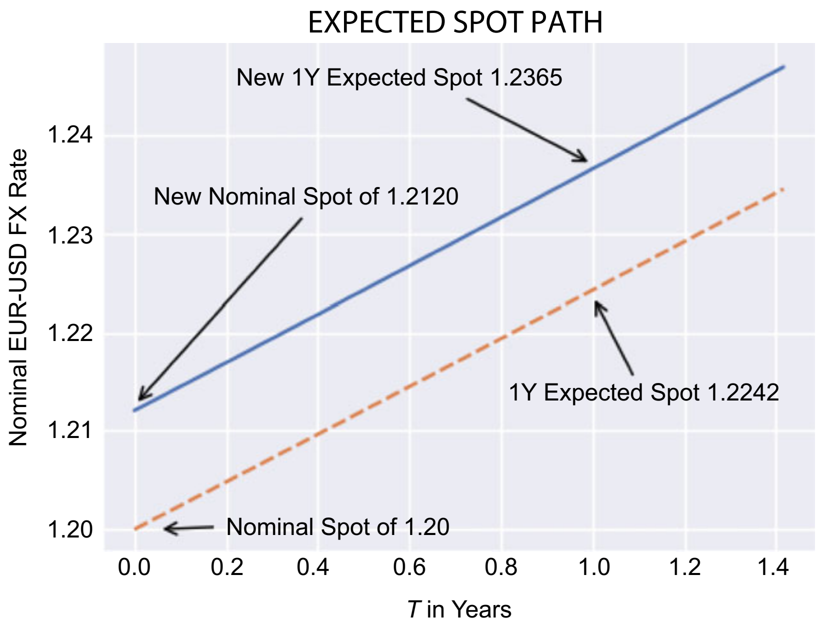


Fig. 3.4 The dashed line shows the initial situation. The nominal FX rate is 1.20. An increase in expected inflation in the U.S. means that the USD is expected to be weaker in the future, and therefore the 1 year expected forward rate must rise. This cheapening in the future value of the USD reflects the expectation that prices in the U.S. will be higher in the future and occurs to keep the 1 year real exchange rate fixed at 1.0. This explains why the solid line is above the dashed line. However, unlike example (i), the U.S. interest rate does not move higher to reflect higher inflation expectations. The CIP relationship tells us that the forward path must not steepen. The nominal EUR-USD FX rate therefore rises

not able to rise to account for the higher expected inflation—r_d remains at 2%. Why might such a situation arise in practice?

Several central banks around the world have implemented Quantitative Easing (QE) programs. This involves the central bank purchasing government bonds in exchange for reserves, and may prevent yield curves from rising. In the case of Japan, such fixed nominal yields have been made explicit through the Bank of Japan (BOJ's) policy of yield curve control (YCC), in which it has explicitly committed to peg the interest rate on 10 year Japanese Government Bonds (JGBs) at around zero per cent. This policy was implemented in 2016, and continues to the time of writing. Another circumstance is if the central bank chooses to remain *behind the curve*, meaning that the central bank does not raise interest rates at the pace of inflation. For example, after the Coronavirus crisis of 2020, inflation returned, but not wanting to risk derailing the recovery, and central banks chose to keep interest rates low.

If the nominal interest rate remains unchanged, then the CIP relationship, Eq. (2.1), implies that the steepness of the forward curve remains unchanged. Therefore, for the 1 year forward to remain at 1.2365, the spot rate must rise today. This is shown by the solid line in Fig. 3.4. Unlike in example (i), a rise in inflation with a fixed nominal interest rate leads to both the spot and forward FX prices moving in response.

A Real Return and Capital Flow Based Derivation

The derivation in this example assumes that PPP holds at the 1 year horizon, and then backs out the move in the spot FX rate to satisfy the CIP relationship. This section arrives at the same result by considering the economic flows that are implemented by risk neutral investors seeking to maximise their real return. The key to understanding these flows is to note the implication of the Fisher equation that rising inflation with a fixed nominal interest rate means decreasing a real interest rate. A falling real interest rate leads investors to seek a better real return abroad and thereby buy EUR vs selling USD.

Again, the starting situation is the dashed line in Fig. 3.4. Market participants then revise their estimate of U.S. inflation higher in response to new economic data. They also understand that the central bank has fixed the nominal interest rate. If the investor keeps her investments in the U.S. then she will realise a real return of -1%. To see why, consider the investor's decision. She may spend 120 USD to consume 1 widget today, or alternatively she may purchase a 120 USD worth of U.S. government bonds. In 1 year, the bonds will return her $e^{0.02\times 1} = 122.40$ USD. However, with $\pi_d^e = 3\%$, the price of a widget is expected to be a further 1% higher at 123.65 USD. She can therefore expect to consume 1% fewer widgets in 1 year in the future than she can today.

However, with EUR-USD spot trading at 1.20, a savvy international investor realises that she can do better than this if she shifts her investment into Europe at a spot rate of anything better than 1.2120. She implements the following strategy. She sells 120 USD of notional of her U.S. bond, and then sells the USD she obtains to buy 100 EUR. She then purchases 100 EUR of a European bond. In 1 year the European bond will return 100 EUR (a 0% interest rate) with certainty. Since inflation in Europe is expected to be 0%, she can expect to consume 1 full European widget in 1 year. The investor can continue increasing her expected real return until she drives the spot rate up from 1.20 to 1.2120.

Once the spot rate has reached 1.2120, there is no further advantage to pursuing the strategy. The reason is that European bonds are now 1% more expensive today, and so she can buy 1% less of them with 120 USD. She will therefore obtain 1% fewer widgets in 1 year from today whether she leaves her investments in U.S. bonds or in European bonds.

However, with EUR-USD spot trading at 1.2120, the CIP relationship implies that the 1 year forward contract trades at 123.65, because $r_f = 0\%$ and $r_d = 2\%$. The investor may therefore enter into the 1 year forward contract today to sell 100 EUR for 123.65 USD. She may then buy 1 widget with these USD in 1 year with certainty. She therefore improves her consumption of widgets by 1% relative to had she kept her investment in the U.S., and she does not need to substitute European widgets for American ones. Once again, we end up with the forward curve shown by the solid line in Fig. 3.4.

Example (ii) Summary

The important take away is that the USD must weaken by the same percentage as the U.S. real interest rate moved lower. In this example, the U.S. real interest rate fell by 1%, and the USD weakened against the EUR by 1%. At this point the investor becomes indifferent between investing in Europe or the U.S.

After the EUR-USD spot rate has repriced, it may superficially remain attractive to save in Europe, rather than in the U.S. because the European real interest rate remains at 0%, compared with a U.S. real interest rate at −1%. However, this is not true with regard to consumption. The real exchange rate today has already moved in favour of Europe via the rise in the nominal EUR-USD spot rate. This means that European assets are now more expensive by 1% and the investor's consumption will not be 1% higher in 1 year, despite the real yield advantage.

The loose heuristic for the purpose of practical trading is that the investor should invest in countries with higher real yields, but only until the nominal exchange rate move fully offsets the difference in real interest rates. At that point, the investor should stop. A higher real interest rate alone is not a reason to invest in a currency.

A final point to note is that while PPP is satisfied (in expectation) at the 1y time point, it is no longer satisfied today. The reason is that the EUR-USD spot rate moves to 1.2120 today. However, since widget prices in Europe and the U.S. are assumed not to change today, the real exchange rate deviates from 1.00 to 1.01, since EUR has become 1% more expensive against USD.

Indeed, for PPP to hold at all time points, real interest rates must be equal across the two economies. This example suggests that FX investors who rebalance their portfolios in response to changes in real interest rates across economies either target consuming directly in the economy with the higher real rate in the future, or believe that the PPP (or another fixed real exchange rate) will be restored in the forward market.

A Note on Exchange Rate Policy

Example (ii) suggests that monetary policies such as QE and YCC may lead to larger reactions in the FX spot market to inflation than if such policies are not implemented. Although rarely made explicit, it is often suspected that monetary authorities implement such policies at least partly with the objective of weakening their currency, and thereby boosting the export economy (Chapter 6 discusses the relationship between exports and the exchange rate). Example (ii) suggests that such a policy is only effective in weakening the FX rate if inflation expectations simultaneously rise. If inflation expectations fall when implementing such a policy, then it may have an exaggerated impact in the opposite direction!

In the early days of QE, it was suspected that the policy would lead to higher inflation, and therefore inflation expectations rose on QE announcements. Note, for example, the dramatic weakening in the EUR in Q1 of 2015 following the ECB's announcement that it would purchase 60 billion EUR of bonds each month. However, the evidence on the relationship between QE and inflation is weak, and the mechanism of action remains poorly understood.[3] Therefore, the impact of such policies on currencies in the future is unclear.

A Note on FX Volatility Term Structure

In example (ii), an inflation shock of 1% leads to the spot rate moving by 1%, but also the forward rising by 1%. This dynamic suggests that implied volatility term structures should be flat. Actions by central banks to pin down nominal yields and thereby decrease their volatility should lead to higher volatility in the FX spot market relative to volatility in the forward market. A term structure of implied volatility that is not sufficiently flat may therefore present a trading opportunity if the trader believes that example (ii) appropriately describes the forthcoming FX price dynamics, all else being equal.

[3] For instance, the *Financial Times* writes, "the Bank of England does not understand its own flagship quantitative easing programme, the central bank's internal independent watchdog concluded…". Source: *Financial Times* (2021).

3.4.3 Example (iii): An Interest Rate Hike from the Central Bank

Once again, consider a starting situation as shown by the dashed line in Fig. 3.5. The Federal Open Markets Committee (FOMC) unexpectedly hikes the interest rate by 1%. How should the EUR-USD FX market react?

An interest rate hike instigated by the central bank to cool the economy is usually accompanied by a fall in inflation expectations. If investors expect that U.S. inflation will be lower, the nominal price of U.S. widgets is expected to be lower than 122.42 USD per widget in 1 year. Let us suppose that it is revised down to 1.2159. To maintain PPP at the 1 year horizon, the 1 year forward price of EUR-USD must therefore fall to 1.2159.

To understand how the spot rate should react, I again apply the CIP relationship. Since r_d is now higher than it was, the forward curve must steepen. This steepening is shown by the solid line in Fig. 3.5.

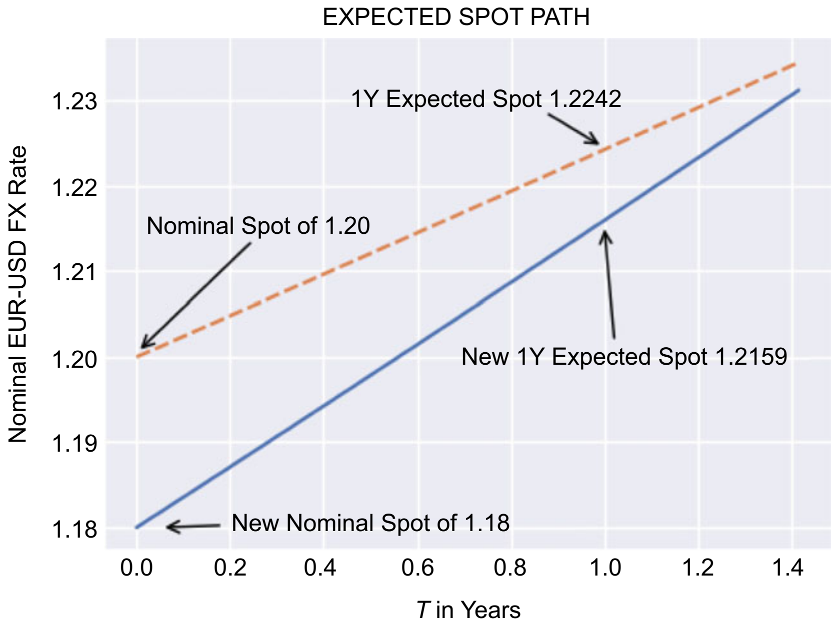


Fig. 3.5 The dashed line shows the initial situation. The nominal FX rate is 1.20. An unexpected increase in U.S. interest rates instigated by the FOMC will, in general, lead to lower inflation expectations in the U.S. For PPP to hold at the 1 year horizon, the EUR-USD forward rate must be lower. However, by the CIP relationship, the forward path must also be steeper. These two effects are additive in terms of the direction in which the spot rate moves. This leads to a large move in the spot FX rate, which is referred to as *exchange rate overshooting*

The large reaction in the spot market is sometimes referred to as *exchange rate overshooting*. The reason is that the movement lower in the 1 year forward is additive to the steepening in the forward path, in the sense that both effects act to lower the EUR-USD spot price. That is, the spot rate falls because the outright level of the forward falls, and then again because the forward path steepens. The new, steep, forward path suggests that the EUR-USD spot price is expected to trend back higher over time, and thereby undo some of the overshoot.

Example (iii) Summary
The important take away is that, just as in case 2, the USD strengthens by the same percentage as the U.S. real yields move higher. To see this, note that r_d increases by 1% (from 2 to 3%) and π_d^e falls to 1.3% (since the expected increase in widget prices is $\ln(1.2159/1.20) = 1.3\%$). Applying the Fisher equation, r^{real} has therefore risen by 1.7%, which is equal to the movement in the EUR-USD FX spot rate, $(\ln(1.18/1.20) = -1.7\%)$. Again, we are able to show that the key driver for the change in the FX spot price is the change in the real interest rate.

3.4.4 Example (iv): FX Response with a Policy Rule

In real markets, examples (i)–(iii) are not independent in that components of some, or all three, may occur at the same time. The FX response is then given by the aggregate impact of each effect. For example, an increase in inflation expectations may trigger a move higher in the nominal interest rate (example (i)), but it may also trigger a response (or expectation of a response) in the direction of an even higher or lower nominal interest rate from the central bank (example (ii)) based on its policy at the time, and this response may cause the real interest rate to move higher, or lower (example (ii)), depending on the size of this response and its impact on inflation expectations. This example considers an FX response that combines the effects discussed in the previous three cases.

Let us begin by understanding the size of a central bank response in normal market conditions. Here, economic theory suggests a natural lower bound; an inflation targeting central bank must hike interest rates by more than the increase in inflation expectations. The Fisher equation tells us that if π_d^e were to increase by 1%, and r_d were to increase by less than 1%, then the r_d^{real} would fall. A lower real interest rate may cause further heating of the economy, and a further increase in π_d^e. Following through this logic, the end outcome is an inflationary spiral. To avoid inflationary spirals, it is therefore crucial that the response is greater than one-for-one. This is known as the

Taylor principle, following the demonstration in Taylor (1993) that it fits the U.S. experience well.

With this in mind, consider the starting situation shown by the dashed line in Fig. 3.6. As before, nominal interest rates are $r_f = 0\%$ in Europe and $r_d = 2\%$ in the U.S, and expected inflation rates are $\pi_f = 0\%$ in Europe and $\pi_d = 2\%$ in the U.S.

A U.S. economic data release causes market participants to revise π_d higher. In the absence of a central bank response, π_d^e would rise by 1%, as it did in examples (i) and (ii). However, the central bank responds, or is expected to respond, with a 1.5% hike, causing r_d to move from 2 to 3.5%. The central bank response also tempers π_d back lower. I assume that it rises by 0.66–2.66%, rather than to 3%. For PPP to be maintained at the 1 year horizon, the 1 year forward must therefore rise by 0.66%, to 1.2324. Then, by the CIP relationship, a higher r_d results in a steeper forward path.

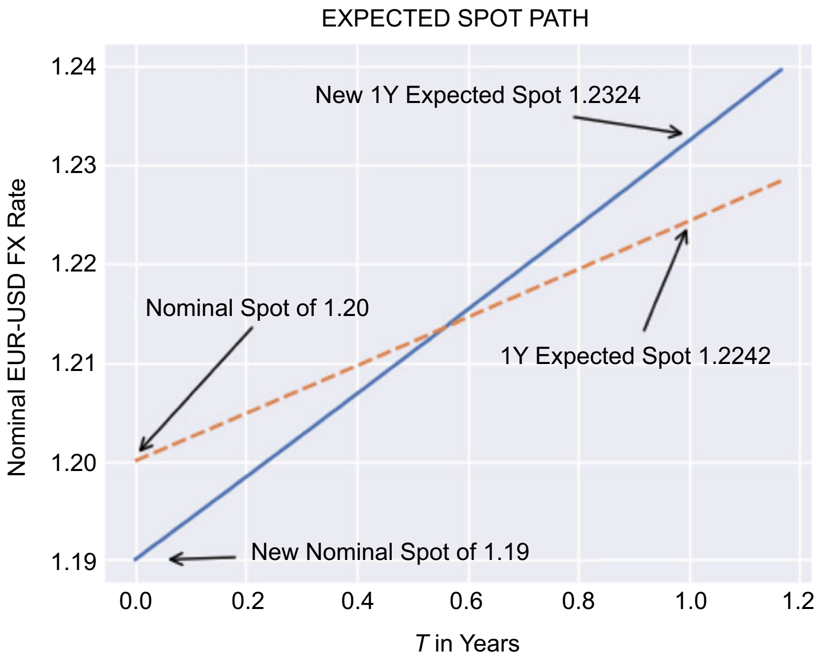


Fig. 3.6 The dashed line shows the initial situation. The nominal FX rate is 1.20. An increase in inflation expectations of 1% leads to a central bank response. The central bank hikes the nominal interest rates. This leads to a tempering of inflation expectations (to 0.66% in this example), and, by CIP, a steeper forward path and a lower FX spot rate. This is shown by the solid line

The spot EUR-USD FX rate therefore falls to 1.19. The new forward path is shown by the solid line in Fig. 3.6.

Example (iv) Summary
Since π_d has risen by 0.66% and r_d by 1.5%, by the Fisher equation r_d^{real} has risen by 0.84%. As in examples (i) to (iii), the nominal FX rate strengthens in favour of the USD by $\ln(1.19/1.20) = -0.84\%$. Again, the key driver for the change in the FX spot price is the change in the real interest rate.

A Note on FX Volatility
Examples (i)–(iv) show that with stable expected future real exchange rates, all volatility in the nominal FX spot rate is driven by changes in the real interest rate. Multi asset volatility investors interested in the relationship between FX volatility and interest rate volatility may note the key take away here; even though interest rates are often seen as a driver of FX rates, higher nominal interest rate volatility does not necessarily translate into higher FX volatility. In the context of the models presented in this chapter so far, and before we account for the effect of changing risk premiums, it will only do so if the higher interest rate volatility is driven by volatility in real FX rates rather than volatility in inflation expectations.

3.4.5 Example (v): FX Response with Changing Risk Premiums

Observers of emerging market FX price dynamics may note that inflation data can cause opposing reactions, depending on the currency involved. For example, while the spot EUR-USD price often falls (the USD strengthens) when positive U.S. inflation data is released (as discussed in example (iv)), this may not be true if we replace the USD with the Turkish Lira (TRY) or several other EM currencies. This leads to the opposite result that EUR-TRY rises (the TRY weakens) when positive Turkish economic data is released. How can we understand this behaviour?

One explanation is that central bank credibility may be lower in EM economies when compared with DM economies. By this, I mean that market participants may doubt the willingness of the central bank to take the tough steps in terms of interest rate hikes that are needed to cool an overheating economy. Such a situation may occur if the central bank lacks de facto independence from the fiscal authority. This may lead to lower real interest rates in the economy and thereby a weakening in the spot FX rate.

However, it is often observed that EM currencies weaken even if the central bank does follow a Taylor rule policy response and raise the real interest rate in response to higher inflation. How can a currency weaken with an increasing real interest rate? The framework presented so far does not allow this; in examples (i)–(iv) the currency has not only moved in the same direction as the change in the real interest rate, but also by the same percentage. To explain this phenomenon, we must include the concept of the risk premium that was absent in examples (i)–(iv).

Risk Premiums

Consider the starting situation shown in the upper section in Fig. 3.7. Note that there are now two lines to describe the starting situation. The dashed line shows the expected path of the spot FX rate, and the solid line shows the forward path. I continue to assume that inflation in Europe is $\pi_f = 0\%$, but I assume a higher rate of inflation in the domestic country, Turkey in this example, at $\pi_d = 10\%$. Applying the fixed expected exchange rate approach, the 1y expected EUR-TRY spot FX rate is therefore 9.95, 10% higher than the FX spot rate of 9.0.

In examples (i)–(iv), I was able to describe the starting situation using just one line because in the absence of a risk premium, the expected spot path and forward path are equal to each other (see Eq. 3.6). Here, I allow a positive risk premium for investors holding a long position in TRY, and hence the forward path is above the expected path of spot. I continue to assume a nominal interest rate on EUR of $r_f = 0\%$, but a nominal interest rate on TRY of $r_d = 15\%$. Applying the CIP relationship, this leads to a 1 year forward price of 10.45. The risk premium, which I denote by r^p, at the 1 year horizon is therefore $r^p = \ln(10.45/9.95) = 5\%$.

Although it has not always been the case, at the time of writing, investors in the U.S. and several other (typically) developed market economies associate deflation with economic turbulence and therefore rising risk premiums, and inflation with economic expansions and falling risk premiums. The opposite is true in, for example, TRY and several other EM currencies, where higher inflation is associated with economic turbulence and a rising risk premium.

Suppose that a positive inflation shock arrives. After incorporating expectations of a central bank response, investors subsequently expect that prices denominated in TRY will be a further 10% higher in 1 year than they had previously anticipated, and therefore revise higher their expectation of the EUR-TRY FX spot rate in 1 year from 9.95 to 11.04. This is shown by the dashed line in the lower section of Fig. 3.7.

The next step is to consider what happens to the 1 year forward. Suppose that the increased risk aversion leads investors to demand a higher risk premium of, say, 15% rather than 5%. The 1 year forward therefore moves to 12.82, since $\ln(12.82/11.04) = 15\%$. With the 1 year forward point pinned

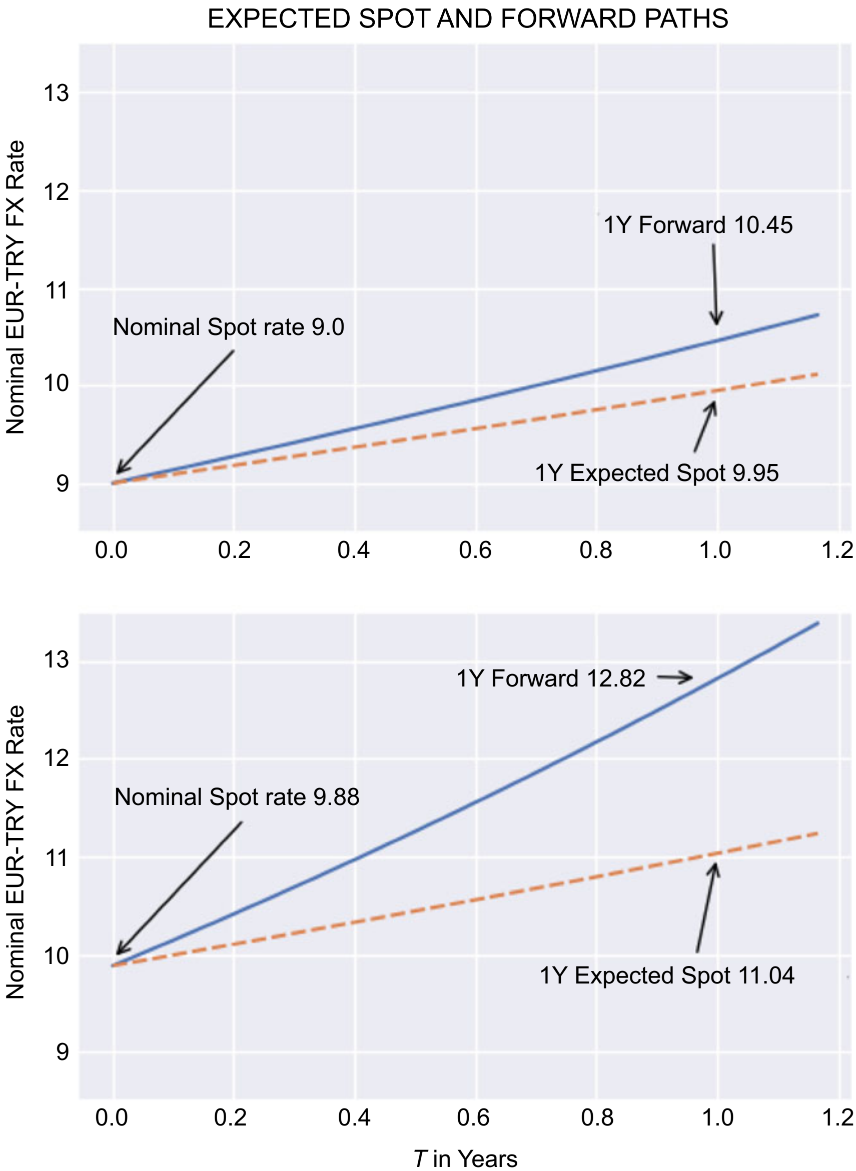


Fig. 3.7 The upper plot shows the starting situation. EUR-TRY is trading at 9.0, and is expected to be 9.95 in 1 year. The risk premium is 5%, leading to a 1 year forward price of 10.45. When an inflation surprise occurs, investors revise their estimate of EUR-TRY spot in 1 year higher to 11.04, but also increase the risk premium they require to hold TRY during the economic turmoil to 15%. The 1 year forward therefore moves to 12.82. Nominal interest rates in TRY increase from $r_d = 15\%$ to $r_d = 26\%$ in anticipation of, or in response to a central bank hike. This steepens the forward curve to dampen the spot reaction. The CIP relationship tells us that the spot price ends at 9.88, higher by 9%

down at 12.82, our final task is to use the CIP relationship to calculate the steepness of the new forward path, and trace this backwards to determine the new EUR-TRY spot FX rate. The larger the extent to which the market prices in interest rate hikes from the central bank, the steeper the forward path, and the smaller move in the spot FX rate.

In Fig. 3.7, I assume that the nominal interest rate moves from $r_d = 15\%$ to $r_d = 26\%$. That is, the market anticipates a central bank response of 11%, which is more than the 10% increase in inflation, as prescribed by the Taylor rule. The solid line shows the new forward path based on the CIP relationship. This results in a spot FX rate of 9.88, an increase of 9% from the starting level of 9.0.

Example (v) Summary

In countries where inflation is associated with economic underperformance, the currency risk premium increases with higher inflation expectations. Therefore, the movement in the forward price is exacerbated. First, the forward moves higher because higher inflation leads to a higher expected future spot FX rate. Second, the wedge between the expected future spot rate and the forward widens because the currency risk premium increases. The nominal spot rate also typically moves higher, but the size of this move is less than that of the forward. The reason is that nominal interest rates typically move higher in anticipation of, or as a result of, a hike from the central bank. This steepens the forward curve and this dampens the spot reaction.

The key takeaway for investors is to note that the additional risk premium must exhibit itself *somewhere*. If the nominal interest rate rises (likely driven by a central bank hike, or anticipated hike) then investors are compensated for holding exposure to the risky currency, TRY in this example, through this channel. However, if the rise in the nominal interest rate is not sufficient to compensate investors for *both* higher inflation, and higher required risk premiums, then the spot value of the currency cheapens to provide the required compensation. In this example, inflation increased by 10% and the required risk premium increased by 10%. The 11% increase in the nominal interest rate therefore left 9% of required compensation that was subsequently provided by a 9% weakening in the TRY FX spot rate.

To summarise, the trading heuristic is that for every percentage point that nominal interest rates do not rise to cover the sum of the change in inflation expectations and the change in the required risk premium, results in a percentage point of weakening in the risky currency.

Finally, note that this principle works both ways. In (typically DM) countries where higher inflation is associated with stronger economic growth, the required currency premium decreases as inflation expectations move higher.

This leads to a dampening, or even a reversal in the spot move. To illustrate this, I consider example (iv) again, but including the effect of risk premiums.

Suppose that the 1 year forward begins $r^p = 1\%$ higher than the 1 year expected spot price, but this risk premium falls to 0.5% in response to the increase in inflation expectations. Then, with a 0.66% increase in π_d and 1.5% increase in r_d, the EUR-USD spot price must fall (USD strengthens) by $-1.5\% + 0.66\% -0.5\% = -1.34\%$. That is, the since higher inflation in the U.S. is associated with better economic outcomes, the spot USD strengthens by more than it did in example (iv) when risk premiums were excluded.

A Note on FX Volatility Term Structure

In example (v), an inflation shock of 10% leads to an increase in the risk premium of 10%. These two effects lead to an increase in the forward of 20%. However, as in example (i), the increase in the nominal interest rate leads to a smaller move in the spot FX price, 9% in this example. This dynamic of the price of the forward behaving in a manner that is substantially more volatile than the spot FX rate suggests that implied volatility term structures should be upward sloping. A term structure of implied volatility that is not sufficiently upward sloping may therefore present a trading opportunity if the trader believes that example (v) appropriately describes the forthcoming FX price dynamics, all else being equal.

3.4.6 The Real Interest Rate and the Risk Premium

The reader may have noted from example (v) that there is a relationship between the real interest rate and the risk premium. In the starting situation (the upper chart in Fig. 3.7), $r_d = 15\%$ and $\pi_d = 10\%$, leaving $r_d^{\text{real}} = 5\%$. This is exactly equal to the risk premium, $r_d^p = 5\%$. Recall, by assumption, $r_f = \pi_d = r_d^p = 0\%$.

The fact that $r_d^{\text{real}} = r_d^p$ is not a coincidence. The CIP relationship imposes that the slope of the forward path is r_d. The fixed expected real exchange rate model imposes that slope of the expected spot path is the rate of inflation, π_d. Therefore, by construction, the gap between the forward and expected spot rate is $r_d - \pi_d$, which is r_d^{real}. The gap between the forward and expected spot rate gap is also the risk premium, as we discussed in Chapter 2.

How does this change after the inflation shock? Since r_d rises to 26% and π_d rises to 20%, r_d^{real} rises to 6%. However, r_d^p rises to 10%. Put another way, of the 10% risk premium on offer to investors, 6% is in the real interest rate, and a further 4% is in the cheapening of the real TRY spot rate relative to its PPP value of 1.0.

The key takeaway for FX investors looking to calculate the risk premium on a currency investment is to note that it is only equal to the real interest rate if the currency is expected to weaken at the rate of inflation. If the spot FX price has already weakened from its expected real future value, then the rate of further depreciation may be slower than inflation, and therefore the risk premium may be higher than the real rate of interest.

To calculate the risk premium available, therefore, the investor must sum the real interest rate and the cheapening in the spot FX rate. In the PPP model this is straight forward; the investor simply calculates the percentage deviation in the real exchange rate from 1.0. However, I discussed reasons why the real exchange rate may diverge from its PPP value of 1.0 in Sect. 3.1.2 even before including risk premiums. Therefore, to calculate the available risk premium in practice, the investors must calculate the deviation of the real exchange rate away from its *normal* value, and add the real exchange rate.

Defining *normal* value is beyond the scope of this chapter. However, a natural starting point is to consider the level of the real exchange rate over past periods.

3.5 FX Price Dynamics—The General Case

Section 3.4 described FX price movements in response to changes in interest rates, expected inflation and risk premiums in the context of fixed expected real exchange rate models. This section formalises the intuition that we developed by deriving an equation to describe the general case.

There are three key equations. The first formalises the meaning of a fixed expected real exchange rate model. To begin, note that,

$$S_t \frac{P_{t,f}}{P_{t,d}} = \mathbb{E}_t^P \left[S_T \frac{P_{T,f}}{P_{T,d}} \right]. \tag{3.7}$$

This equation states that the expected real exchange rate at time T in the future (the right-hand side), is equal to the real exchange rate today (the left-hand side). This is simply a statement of Eq. (3.1) at time points t and T.

To simplify things further, I assume that the inflation rate that will be realised is not random, but known to be π_d and π_f with certainty. Then, Eq. (3.3) becomes,

$$P_{T,d} = P_{t,d} e^{\pi_d \tau} \text{ and } P_{T,f} = P_{t,f} e^{\pi_f \tau}. \tag{3.8}$$

Substituting Eq. (3.8) into Eq. (3.7) gives us our first key equation,

$$\mathbb{E}_t^P[S_T] = S_t e^{(\pi_d - \pi_f)\tau}. \tag{3.9}$$

This equation states that investors expect the FX spot rate to move in line with the inflation difference between the FOR and DOM currencies.

The second equation that we require is inspired by the currency risk premium model of Eq. (2.6) of Chapter 2, which states that $f_t(T)$ is equal to $\mathbb{E}_t^P[S_T]$ plus a risk premium. I re-express this equation as,

$$f_t(T) = \mathbb{E}_t^P[S_T] e^{\left(r_d^p - r_f^p\right)\tau}. \tag{3.10}$$

Here, r_d^p and r_f^p refer to the risk premiums in the DOM and FOR currencies respectively. For instance, in example (v) of Sect. 3.4.5 involving EUR-TRY, the starting situation had $f_t(T)$ higher than $\mathbb{E}_t^P[S_T]$ by 5% at the 1 year horizon. Therefore, $\left(r_d^p - r_f^p\right) = 5\%$.

Substituting Eq. (3.9) into Eq. (3.10) gives

$$f_t(T) = S_t e^{\left((\pi_d - \pi_f) + \left(r_d^p - r_f^p\right)\right)\tau}. \tag{3.11}$$

In words, this equation says that the forward price is higher than the spot price by the difference in inflation in the DOM and FOR currencies, plus the difference in their respective risk premiums. In the starting situation in example (v) the forward was 15% above the spot rate of which 10% arose from the inflation differential and 5% from the risk premium.

The third and final equation that we require is the CIP relationship, Eq. (2.1). I rearrange it here, making S_t its subject.

$$S_t = f_t(T) e^{-(r_d - r_f)\tau}. \tag{3.12}$$

With these equations in place we are ready to derive the general equation for FX price changes.

The starting situation is obtained by substituting Eq. (3.12) into Eq. (3.11). This results in the following relationship,

$$r_d^p - r_f^p = (r_d - r_f) - (\pi_d - \pi_f). \tag{3.13}$$

This equation tells us that when the spot rate is expected to depreciate by the inflation rate (as in the PPP and other fixed expected exchange rate models), then the risk premium differential is equal to the real exchange rate differential. I discussed this idea in the context of example (v) in Sect. 3.4.6.

Next, economic news arrives causing market participants to revise π_d and π_f to new values, π_d^* and π_f^*. The central bank reaction (or the market's anticipation of a central bank reaction) causes r_d and r_f to take new values r_d^* and r_f^*. Based on Eq. (3.11), the forward moves to a new value, which I denote by $f_t^*(T)$, and is given by

$$f_t^*(T) = S_t e^{\left((\pi_d^* - \pi_f^*) + \left(r_d^{*p} - r_f^{*p}\right)\right)\tau}. \tag{3.14}$$

Finally, I apply Eq. (3.12) to calculate the new value of the spot FX rate, which I denote by S_t^*,

$$\begin{aligned}
S_t^* =& f_t^*(T) e^{-(r_d^* - r_f^*)\tau} \\
=& S_t e^{\left((\pi_d^* - \pi_f^*) + \left(r_d^{*p} - r_f^{*p}\right) - (r_d^* - r_f^*)\right)\tau} \\
=& S_t e^{\left((\pi_d^* - \pi_f^*) + \left((r_d^p + \Delta r_d^p) - (r_f^p + \Delta r_f^p)\right) - (r_d^* - r_f^*)\right)\tau} \\
=& S_t e^{\left((\pi_d^* - \pi_f^*) + \left((r_d - \pi^d + \Delta r_d^p) - (r_f - \pi_f - \Delta r_f^p)\right) - (r_d^* - r_f^*)\right)\tau} \\
=& S_t e^{\left((\Delta\pi_d - \Delta\pi_f) + \left(\Delta r_d^p - \Delta r_f^p\right) - (\Delta r_d - \Delta r_f)\right)\tau} \\
=& S_t e^{\left((\Delta r_f^{\text{real}} - \Delta r_d^{\text{real}}) + \left(\Delta r_d^p - \Delta r_f^p\right)\right)\tau},
\end{aligned} \tag{3.15}$$

where in the third line I substitute in $r_d^{*p} = r_d^p + \Delta r_d^p$ and $r_f^{*p} = r_f^p + \Delta r_f^p$, in the fourth line I substitute in Eq. (3.13), and in the fifth line I substitute in the Fisher Equation, Eq. (3.4). Finally, taking logs of Eq. (3.15) gives us the required result,

$$\ln \frac{S_t^*}{S_t} = \left((\Delta r_f^{\text{real}} - \Delta r_d^{\text{real}}) + \left(\Delta r_d^p - \Delta r_f^p\right)\right)\tau. \tag{3.16}$$

In example (v), Δr_f^{real} and Δr_f^p were both 0%, because I assumed that there were no changes on the EUR side. However, $\Delta r_d^{\text{real}} = 1\%$ and $\Delta r_d^p = 10\%$, which resulted in the cheapening of the TRY spot price of 9% against the EUR.

3.6 Equity Investments and FX-Equity Correlation

Section 2.3.3 discussed the analogy between equity and FX investments. This section extends the ideas and methods discussed in this chapter to equity investments through two examples. Example (vi) studies a *taper tantrum*, a situation where the central bank signals a higher nominal interest rate at a time when the economy does not anticipate higher inflation. This causes an increase in the risk premium, a fall in equity prices, and a rise in the USD against risky currencies. Example (vii) studies a so-called *risk off* event, a situation whereby negative economic news permeates the economy and the nominal interest rate falls as the market anticipates that the central bank will cut. The risk premium rises leading to a fall in equity prices, a rise in the USD against risky currencies, and a fall in the USD against the safe-haven currencies.

The intention here is to understand movements in equity prices, but also an important source correlation between FX and equity prices, which arises fact that equity prices are denominated in a currency and that both currencies and equity prices are subject to changes in risk premiums. This idea will become clear through the examples.

3.6.1 Example (vi): Taper Tantrum

Let us begin with the situation shown in the upper section of Fig. 3.8. The S&P 500 is trading at 3700. The dashed line shows that investors expect that this price will grow at 4% per year, and will therefore be 3851 in 1 year from today. The solid line shows the forward path, derived from the equity equivalent of the CIP relationship discussed in Sect. 2.3.3. Here, I assume that $r_d = 1\%$, and the dividend yield, which I denote by r_{div} is 3%. The forward path is given by $f_t(T) = \exp(r_d - r_{div})\tau$ and the 1 year forward is therefore at 3627, 2% below the spot price. Note also that the 1 year forward is $\ln(3267/3851) = 6\%$ below the 1 year expected spot price. This is the well documented equity risk premium (see, for example, Chap. 21 of Cochrane (2005) for a review).

Next, the central bank signals that it may raise the nominal interest rate at a time when investors assess that the economy is not ready for such a move in that they do not see inflationary or other overheating pressures. The reaction in the equity market is shown in the lower section of Fig. 3.8.

First, investors lower their expectation of where the S&P 500 will be in 1 year. The reason is that a higher interest rate lowers inflation expectations.

If the prices of goods sold are expected to be lower, then dividends (denominated in USD) are expected to be lower also. This deflationary impact on the 1 year expectation of the S&P 500 is analogous to how the 1 year expected forward in EUR-USD fell in response to an interest rate hike in example (iii); we have simply replaced EUR-USD with S&P500-USD.

In the case of the equity market there may be a further additive impact. The higher interest rate may negatively impact investment, and further lower future dividends. This may decrease investor expectations of the 1 year S&P 500 spot level even further. In the figure, I assume that the 1y expected value of the S&P 500 falls by 1% to 3813.

Next, I assume a rise in the required risk premium rises as investors become fearful of the potentially negative impact on consumption of the central bank's move. In Fig. 3.8, the risk premium rises from 6 to 8% putting the 1y forward at 3520.

Finally, I apply the equity market equivalent of the CIP relationship to trace back the impact on the spot market from the fall in the forward. Since the nominal interest rate is higher, and the dividend yield does not move higher, the forward path is less downward sloping than in the initial situation. This leads to a further fall in the spot S&P 500. In Fig. 3.8, the fall is 4%. This can be decomposed as 1% from the revision lower in the 1 year expected price of the S&P 500, 2% from the increase in the risk premium, and a further 1% from the flattening in the forward path.

Equity-FX Correlation

Example (vi) provides some intuition around the two sources of correlation between FX and equities discussed in this chapter, namely inflation, the nominal interest rate and the risk premium.

The correlation that arises from inflation can be summarised as follows. The fact that the S&P 500 is denominated in USD means that lower expected inflation in the U.S. lowers the expected future price of the S&P 500 in USD, and also lowers the expected future price of FOR-USD FX rates.[4]

The correlation that arises from the nominal interest rate arises from the fact that a change in the U.S. nominal interest rate shifts the forward curve in the same direction for the S&P 500-USD and the FOR-USD FX rate.

The correlation that arises from the risk premium is more complicated. Whether this component creates a positive or negative correlation with the equity price depends on whether the currency is *risky*, by which I mean

[4] There may be exceptions for some global stock markets. For instance, the FTSE 100 of the U.K. consists of international companies with substantial foreign earnings. That is, these companies sell only a small proportion of goods in GBP. Therefore, lower expected inflation in the U.K. need not lower expected dividends that are denominated in GBP.

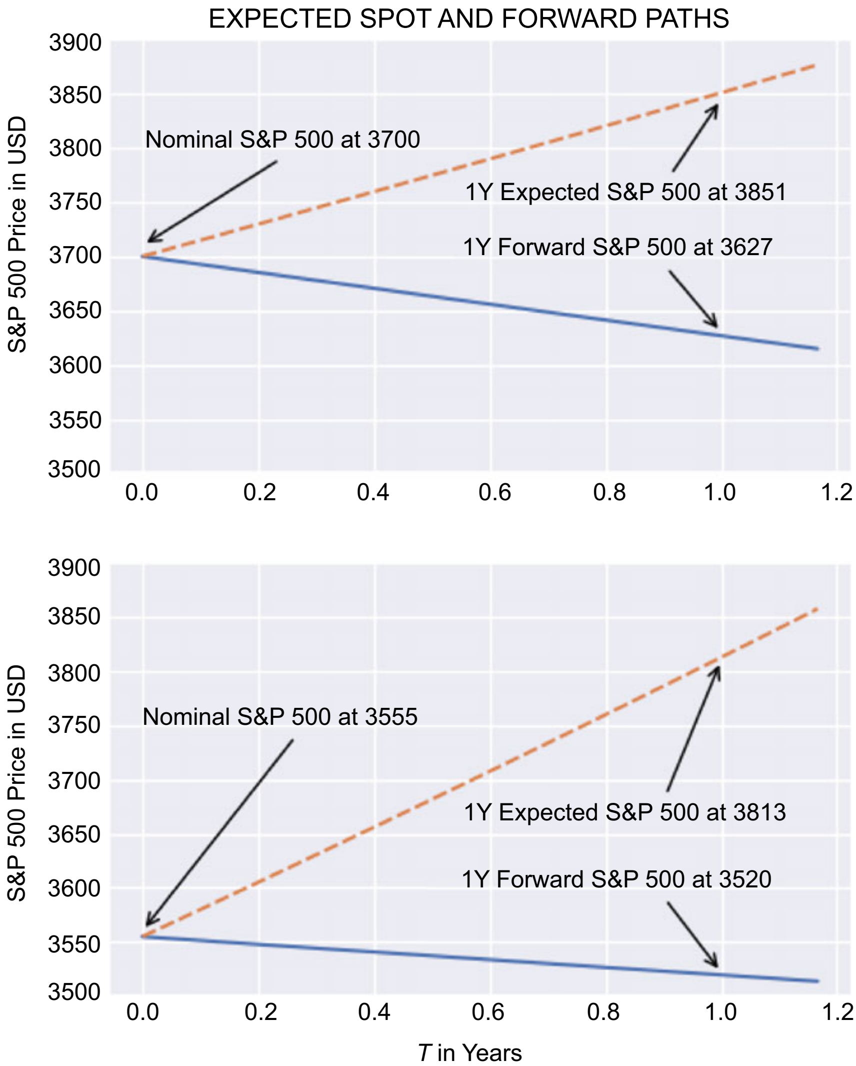


Fig. 3.8 The upper figure shows the initial situation. The dashed line shows that investors expect a rise in the S&P 500 of $\ln(3851/3700) = 4\%$. The solid line shows the forward path, calculated using the equivalent of the CIP relationship with $r_d = 1\%$ and $r_{div} = 3\%$. The lower figure shows the situation following a signal of an interest rate hike by the central bank. The dashed line shows that investors expect a lower future value for the S&P 500 by 1%. The larger gap between the dashed and solid lines shows that investors also demand a higher risk premium of $\ln(3813/3250) = 8\%$. Finally, a flatter forward path accounts for the rise in the U.S. nominal interest rate. The S&P 500 falls by 4% from 3700 to 3550

one that carries a positive risk premium, such as AUD, MXN or other risky DM or EM currencies, or a *safe-haven* currency that carries a negative risk premium, such as CHF or JPY.

The dynamics associated with risky currencies are shown in Fig. 3.9. There, I take MXN-USD as an example.[5] Note that the dynamics are almost identical to Fig. 3.8. There is a strong correlation between MXN-USD and the S&P 500 stemming from the fact that an increase in the risk premium weakens MXN-USD and the S&P 500, and an increase in the U.S. interest rate also weakens the MXN-USD FX rate and the S&P 500. In the example, the fall in U.S. inflation expectations causes a fall of 1%, the increase in risk premium causes a fall of a further 2% and the flattening of the forward path causes another 1%. All three effects act in the same direction, resulting in MXN-USD falling by 4%.

Next, consider how safe-haven currencies such as the JPY or CHF may behave in this circumstance. The dynamics associated with the CHF-USD FX rate are shown in Fig. 3.10. In the example, the fall in U.S. inflation expectations causes a fall of 1%, the increase in the risk premium causes a rise of 0.5% and the flattening of the forward path causes a fall of 1%. The risk premium effect acts in the opposite direction to the remaining two, and so while the move in CHF-USD nets to lower by 1.5%, the size is smaller than MXN-USD. If the increase in the required risk premium were large enough, then CHF-USD could even move higher!

A Note on FX Volatility

In this example, the impact on the FOR-USD exchange rate of a decrease in inflation expectations in the U.S. and an increase in the risk premium is additive if FOR is a risky currency, carrying a positive risk premium, but offsetting if the FOR currency is a safe-haven currency carrying a negative risk premium. That is, the lower U.S. inflation expectations and a higher U.S. interest rate act to weaken MXN-USD, and a higher risk premium acts to weaken MXN-USD further However, while lower U.S. inflation expectations and a higher U.S. interest rate act to weaken CHF-USD, a higher risk premium acts to strengthen it. Therefore, all else being equal, FX volatility traders should expect volatility to be higher in currencies such as AUD-USD or MXN-USD than in, say, JPY-USD or CHF-USD if the context of an event of the nature of example (vi).

[5] I continue to quote FOR-USD in order to keep the USD on the same side of the quotation, for simplicity and consistency. Market quotation convention is USD-MXN and USD-CHF.

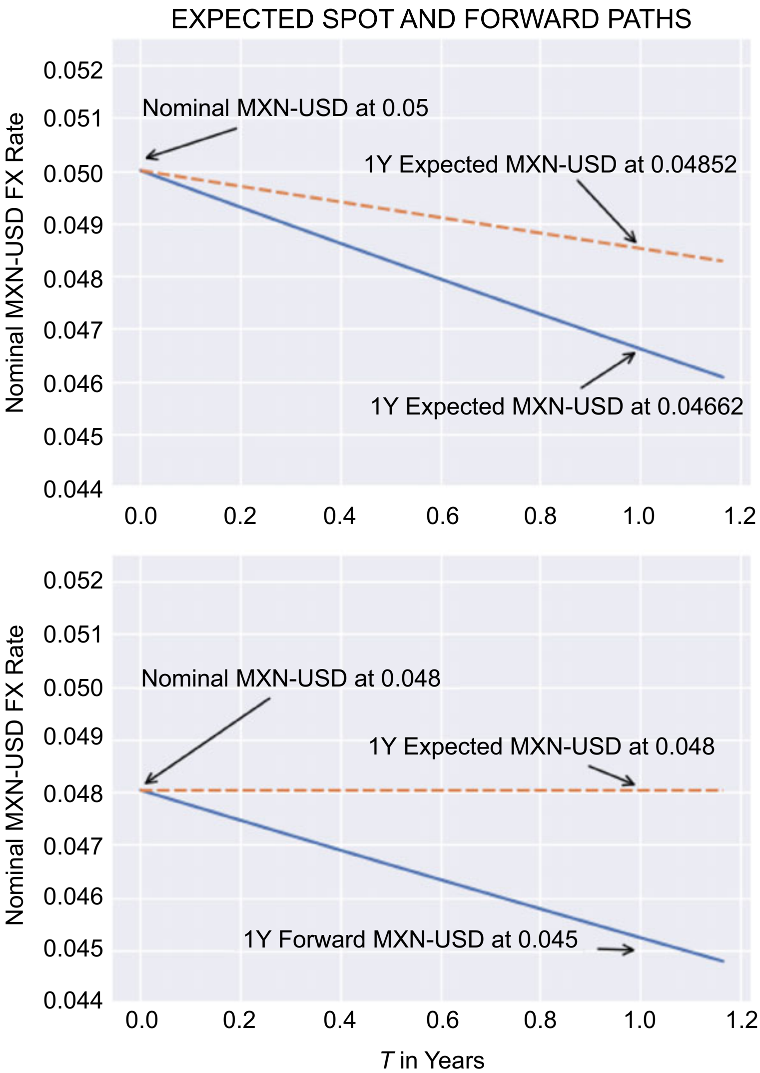


Fig. 3.9 The upper figure shows the initial situation. The dashed line shows that investors expect a fall in the value of MXN-USD of $\ln(0.05/0.0485) = 3\%$ over the year. The solid line shows the forward path, calculated using the CIP relationship with $r_d = 1\%$ and $r_f = 8\%$. The risk premium for holding MXN is $\ln(0.485/0.466) = 4\%$. The lower figure shows the situation following a signal of an interest rate hike by the central bank. The dashed line shows that investors expect a 1% lower future value for MXN-USD of 0.048. The larger gap between the dashed and solid lines shows that investors also demand a higher risk premium of $\ln(0.048/0.045) = 6\%$. Finally, a flatter forward path accounts for the rise in the U.S. nominal interest rate. The MXN-USD FX rate falls by $\ln(0.05/0.048) = 4\%$

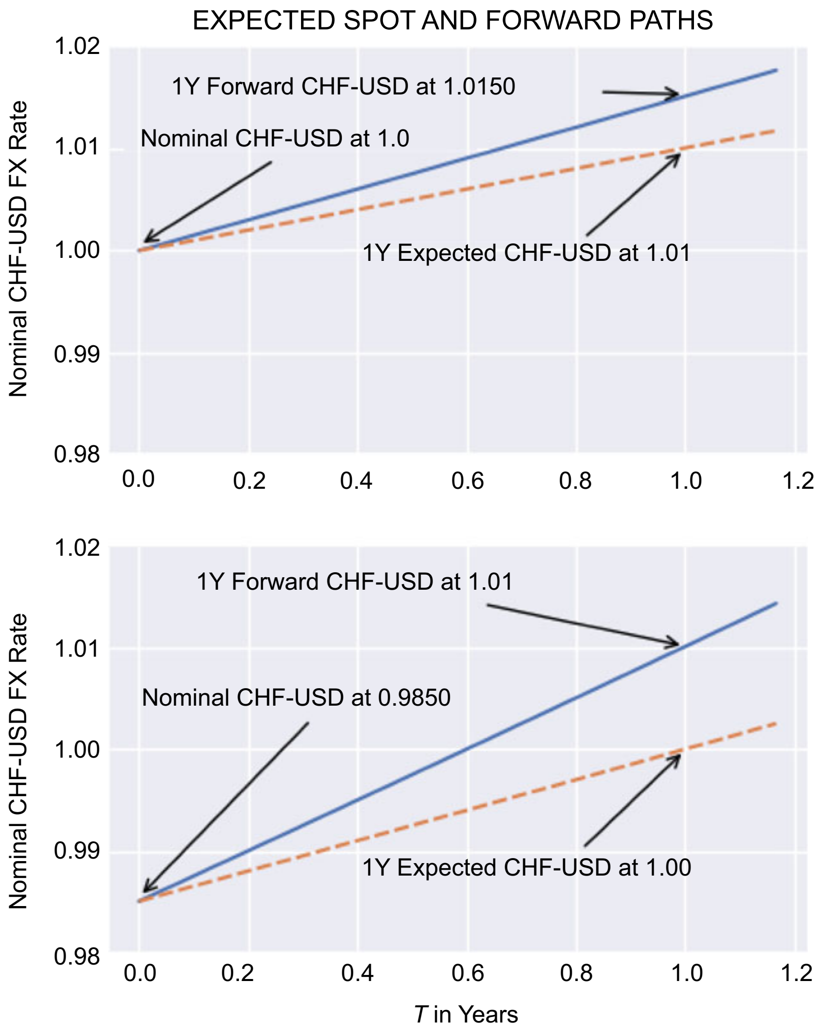


Fig. 3.10 The upper figure shows the initial situation. The dashed line shows that investors expect a rise in the value of CHF-USD over the year of $\ln(1.01/1.00) = 1\%$. The solid line shows the forward path, calculated using the CIP relationship with $r_d = 1\%$ and $r_f = -0.5\%\%$. The risk premium for holding CHF is $\ln(1.01/1.0150) = -0.5\%$. The lower figure shows the situation following a signal of an interest rate hike by the central bank. The dashed line shows that investors expect a 1% lower future value of CHF-USD of 1.0. The larger gap between the dashed and solid lines shows that investors also require a larger (more negative) risk premium of $\ln(1.00/1.01) = -1\%$. Finally, a steeper forward path accounts for the rise in the U.S. nominal interest rate. The CHF-USD FX rate falls by $\ln(1.00/0.9850) = 1.5\%$

3.6.2 Example (vii): An Archetypal "Risk-Off"

Consider a circumstance in which news breaks that suggests that economic growth will fall. The news of the impending Coronavirus pandemic in February 2020, or the turmoil in the U.S. housing market in late 2007 are two particularly notable examples among many. Such environments are typically characterised by falling nominal yields, as markets anticipate both lower inflation and a dovish monetary policy response from the central bank aimed at supporting the economy, falling equity prices driven by falling growth expectations and increasing risk premiums, and falling commodity prices due to falling demand. Market participants commonly refer to such a circumstance as a *risk-off*.

Figure 3.11 shows how the S&P 500 may react. As before, the starting situation has the S&P 500 trading at 3700, with a 4% expected growth rate to 3851 in 1 year. I assume $r_d = 1\%$ and $r_{div} = 3\%$, which results in a 1 year forward of 3627. The risk premium is therefore 6%.

As the negative economic news hits, investors revise down their expectation for the S&P 500 in 1 year. Suppose that this revision is −2%. However, investors also expect that the central bank will cut the nominal interest rate as part of the monetary response to stimulate demand. This will generate higher inflation that would otherwise have occurred had they not taken such action, which raises the nominal value of dividends relative to what they would have been. In the figure I assume that the net impact is −1%. The lower figure shows that the S&P 500 is expected to be 3813 in 1 year.

Next, I assume that the risk premium doubles. The 1 year forward is therefore 12% below 3813, at 3382. Finally, the forward path steepens as a result of the lower nominal interest rate. The S&P 500 falls by 6% to 3485. This fall can be broken down as −6% due to the rising risk premium, −1% due to lower expectations of the S&P500 in 1 year, and +1% due to the lower U.S. nominal interest rate steepening the forward path.

Let us now consider how currencies behave. Again, I take the example of a risky currency, MXN, and a safe-haven currency, CHF. Figure 3.12 shows that MXN-USD falls by 4%, which is broken down as 4% due to a doubling in the risk premium, −2% due to weakening in Mexico's terms of trade due to the fall in commodity prices, +1% due to more inflation in the U.S. than would otherwise have occurred due to the cut (or anticipated cut) in the nominal interest rate, and +1% due a steepening in the forward path.

Finally, Fig. 3.13 shows that CHF-USD rises by 2.5%. This rise can be broken down as +0.5% due to a higher risk premium acting to strengthen the CHF, +1% due to higher inflation in the U.S. relative to Switzerland

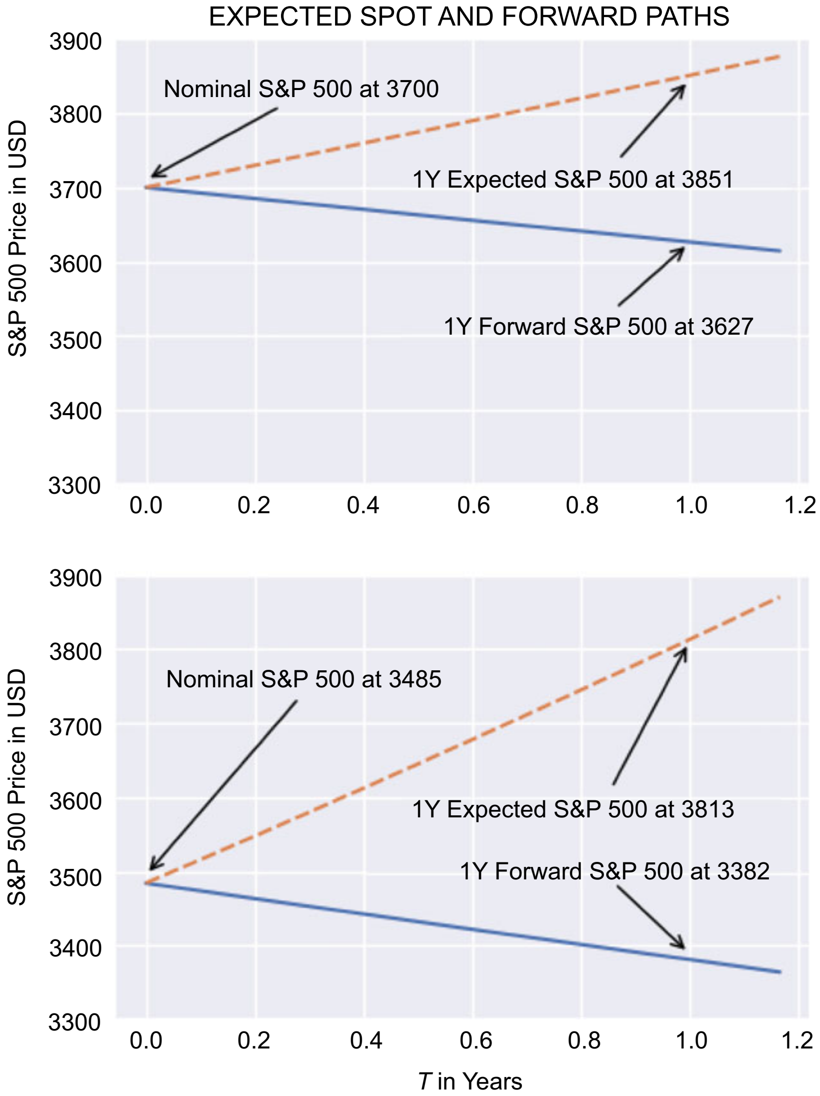


Fig. 3.11 The upper figure shows the initial situation. The dashed line shows that investors expect a rise in the S&P 500 of $\ln(3851/3700) = 4\%$. The solid line shows the forward path, calculated using the equivalent of the CIP relationship with $r_d = 1\%$ and $r_{div} = 3\%$. The risk premium for holding the S&P 500 is $\ln(3851/3627) = 6\%$. The lower figure shows the situation following the negative economic news. The dashed line shows that investors expect a lower future value for the S&P 500 by 1%. The larger gap between the dashed and solid lines shows that investors also demand a higher risk premium of $\ln(3812/3382) = 12\%$. Finally, a steeper forward path accounts for the fall in the U.S. nominal interest rate. The S&P 500 falls by 6% from 3700 to 3485

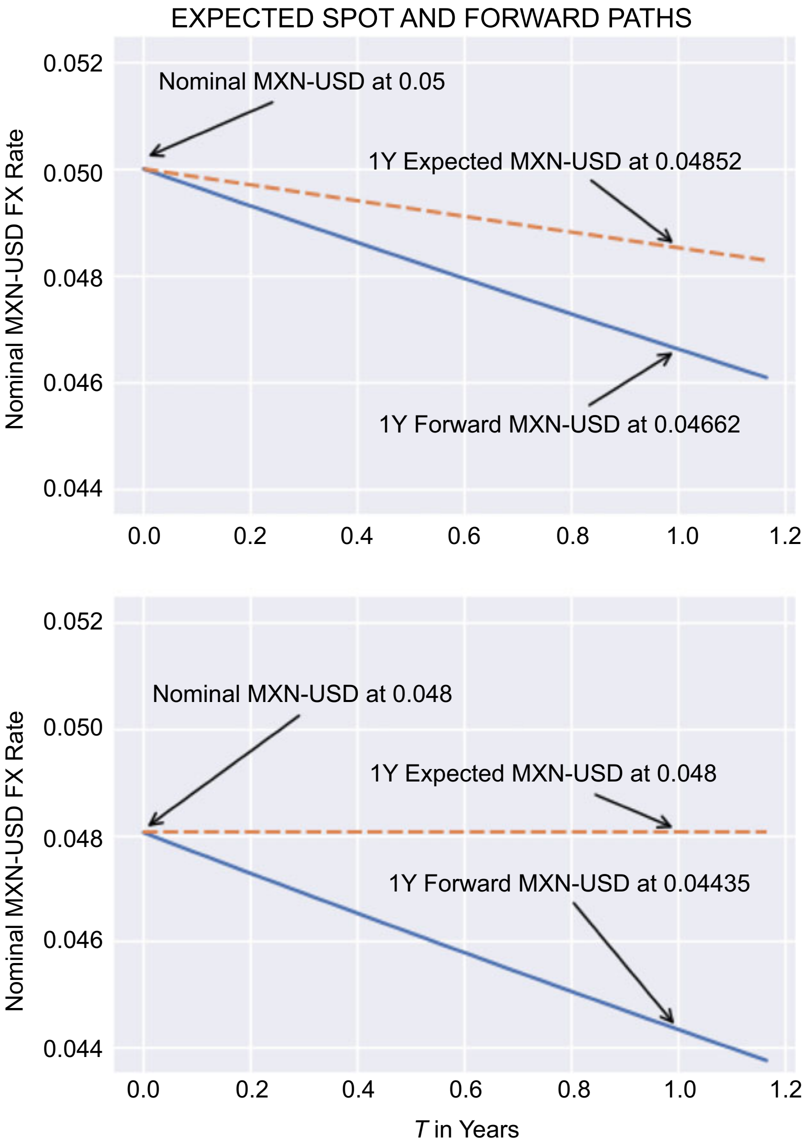


Fig. 3.12 The upper figure shows the initial situation. The dashed line shows that investors expect a fall in the value of MXN-USD of $\ln(0.05/0.0485) = 3\%$ over the year. The solid line shows the forward path, calculated using the CIP relationship with $r_d = 1\%$ and $r_f = 8\%$. The risk premium for holding MXN is $\ln(0.485/0.466) = 4\%$. The lower figure shows the situation following the negative economic news. The dashed line shows that investors expect a 1% lower future value for MXN-USD of 0.048. The larger gap between the dashed and solid lines shows that investors also demand a higher risk premium of $\ln(0.048/0.04435) = 8\%$. Finally, a steeper forward path accounts for the fall in the U.S. nominal interest rate. The MXN-USD FX rate falls by $\ln(0.05/0.048) = 4\%$

following the cut (or anticipated cut) in the U.S. nominal interest rate, and +1% due to the flattening in the forward path.

Note that while S&P-500 and MXN both weakened, the CHF strengthened against the USD. Note also that the marginal impact due to the action (or anticipated action) of the central bank is to decrease the amount by which the S&P and MXN weaken, but to increase the amount that USD-CHF strengthens. That is, the central bank acts to diminish the volatility of risky assets such as stocks and risky currencies, but exacerbates the volatility of safe assets.

Although beyond the scope of this chapter, other central bank action, such as QE or the purchasing of risky corporate debt further acts to decrease the volatility of risk assets through two channels. First, the direct buying of such assets may act to raise their price. Second, such buying may decrease the risk premium relative to what it may be without the presence of the central bank, which again supports risky assets.

A Note on FX Volatility

In the archetypal "risk off" example, the CHF strengthens against the USD because the risk premium effect acts to increase the spot CHF-USD. However, the MXN-USD exchange rate weakens because the risk premium effect acts to weaken the spot MXN-USD, and this effect is larger than the effects that weaken the USD, such as the steeper forward path, and higher U.S. inflation expectations. Therefore, the correlation of USD-MXN and USD-CHF is negative, and the volatility of the cross pairs, CHF-MXN is large. Other risky currency—safe currency pairs, such as AUD-JPY, will also exhibit higher realised volatility in such circumstances through this mechanism. This is also a reason that investors sometimes choose options on such crosses as portfolio hedges against long equity portfolios. It is important to note that based on the analysis presented in this chapter, such hedges may perform in an *archetypal risk off event*, but may not do so in a *taper tantrum*. Recall, in the taper tantrum, CHF-USD and MXN-USD both fell. Therefore, the volatility realised in CHF-MXN is dampened by this positive correlation.

A further point to note relates to macroeconomic correlations. In an increasingly globalised economy, negative macroeconomic news is likely to impact many economies. Therefore, when the FOMC cuts (or is anticipated to cut) the interest rate, so is the ECB, BOE, BOJ and so on. This coordination among central bank impacts FX volatility lower. This is perhaps one of the reasons that FX volatility tends to increase less than volatility in equities or interest rates during global economic downturns.

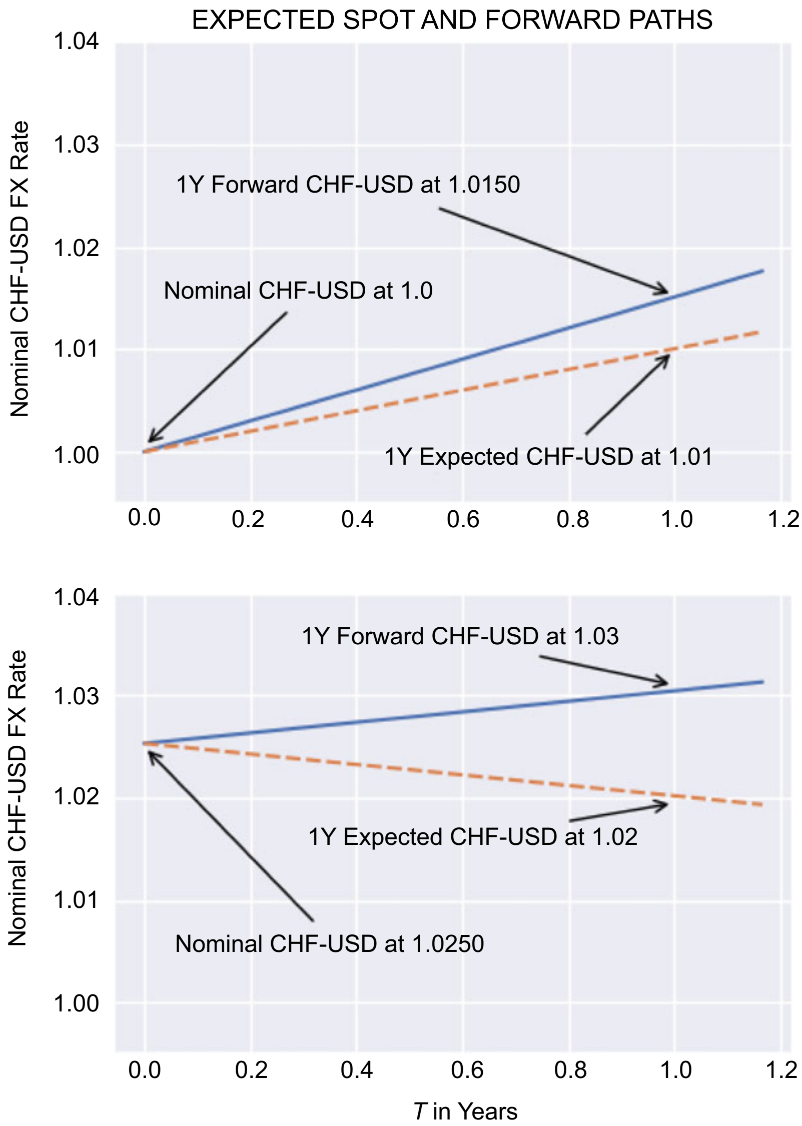


Fig. 3.13 The upper figure shows the initial situation. The dashed line shows that investors expect a rise in the value of CHF-USD over the year of $\ln(1.01/1.00) = 1\%$. The solid line shows the forward path, calculated using the CIP relationship with $r_d = 1\%$ and $r_f = -0.5\%\%$. The risk premium for holding CHF is $\ln(1.01/1.0150) = -0.5\%$. The risk premium is negative because CHF is a safe asset. The lower figure shows the situation following the negative economic news. The dashed line shows that investors expect a 1% higher future value of CHF-USD of 1.02. The larger gap between the dashed and solid lines shows that investors also require a larger (more negative) risk premium of $\ln(1.02/1.03) = -1\%$. Finally, a flatter forward path accounts for the fall in the U.S. nominal interest rate. The CHF-USD FX rate rises by $\ln(1.0250/1.00) = 2.5\%$

3.7 Chapter Summary

- The real exchange rate compares the prices of real goods priced in different currencies.
- The PPP model fixes the real exchange rate to 1.0. Frictions in trading goods across borders lead to the potential for PPP violations. The Balassa-Samuelson theory argues that the violations are due to the presence of non-tradeable goods and productivity differentials across wealthy and poorer nations.
- The fixed expected real exchange rate approach is similar to PPP in that it assumes that the FX rate is expected to weaken at the inflation differential between the FOR and DOM currencies. PPP is a special case of this model, with the real exchange rate set to 1.
- Fixed expected real exchange rate models combined with the CIP relationship allow us to understand the relationship between the spot FX rate, inflation and interest rates.
- In the absence of a risk premium, changes in the real interest rate drive movements in the FX spot rate. A higher real interest rate causes the currency to strengthen.
- In the presence of a risk premium, the sum of the change in the risk premium required by investors, and the real exchange rate drive movements in the FX spot rate. A higher real interest rate causes the currency to strengthen, and a higher required risk premium causes the currency to weaken.
- Correlation in equity markets and FX markets derives from two sources. Changes in inflation expectations typically effect equity markets and FX markets in the same direction, leading to a positive correlation. Changes in the required risk premium may drive the FOR-USD in the same direction as the equity market in the case of risky currencies, or in the opposite direction in the case of safe-haven currencies.

References

Balassa, B. (1964). The purchasing power parity doctrine: A reappraisal. *Journal of Political Economy, 72*, 584.

Financial Times. (2021). Boe criticised by internal watchdog over easing programme. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm

Samuelson, P. (1964). Theoretical notes on trade problems. *Review of Economics and Statistics, 46*, 145–154.

Taylor, J. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*, 195–214.

4

FX Volatility and FX Options

FX options are an important instrument of FX trading. While FX spot and forward contracts, discussed over the course of the previous chapters, provide investors with exposure to the direction in which a currency pair will move, and to interest rates, options provide exposure to *how much* a currency pair will move. More precisely, options provide exposure to a currency pair's future volatility. The first part of this chapter introduces FX options and studies several useful heuristics for practical trading.

This chapter also studies the important idea that it is precisely because volatility is itself uncertain that options exist at all. Even if we do not know the path that the FX spot rate will take in the future, an option is a redundant security if we know the volatility that spot will exhibit because it is replicable through a process called *delta hedging*. We already saw an option replication in a one period model in Example (iii) of Sect. 1.1.3. This argument was formalised in continuous time via the seminal work on options pricing and dynamic replication of Black and Scholes (1973) and Merton (1973). We will see that, arguably the tremendous insight gained from the BSM model is that it is precisely because of the feature of markets that their model does not capture, that volatility is itself uncertain, that options exist at all! In the language of financial economics, options are said to complete the market for volatility risk, because without them, the risk that market volatility rises or falls cannot be hedged. In this chapter, I extend the example from Sect. 1.1.3 to provide the intuition underlying this argument.

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_4

Next, since FX options provide exposure to future volatility, investors can extract a forward looking, or implied, probability distribution from their prices. This distribution provides an investor with information about the range of outcomes that FX options market participants expect a currency pair to have over forward looking time horizons from overnight to, in some cases, up to 20 years. The investor is also able to extract information about the correlations between different currencies over multiple time horizons. The FX options market is deep and liquid with approximately 300 billion USD of notional of options traded each day,[1] and therefore contains information content that is amalgamated from many trades. Extracting this information may prove valuable for forming FX views, for understanding which information is and is not incorporated into the current currency price, for optimal portfolio formation, and for risk management tasks such as calculating Value-at-Risk (VAR), among others. I provide the reader with an intuitive approach to understanding how to extract such information.

The second part of this chapter focusses on volatility in the real FX rate. While Chapter 3 assumed a fixed real FX rate, and discussed changes in interest rates, inflation, and the risk premium as sources of volatility in the nominal FX rate, this chapter studies sources of volatility in the real FX rate itself. The motivation for studying the behaviour of the nominal FX rate in the context of a fixed real FX rate came from the PPP model. If the international arbitrage in real goods, discussed in Sect. 3.1.2, cannot occur due to transport or other frictions, then there is no mechanism to force the real FX rate to remain stable. In Sect. 4.5 I apply the central pricing equation of Chapter 1, Eq. (1.10), to foreign and domestic assets. The resulting equation provides insight into how imperfect macroeconomic risk sharing between consumers in different countries can result in volatility in the real exchange rate. Readers more interested in the economics of real FX volatility than in FX option mechanics may skip directly to Sect. 4.5.

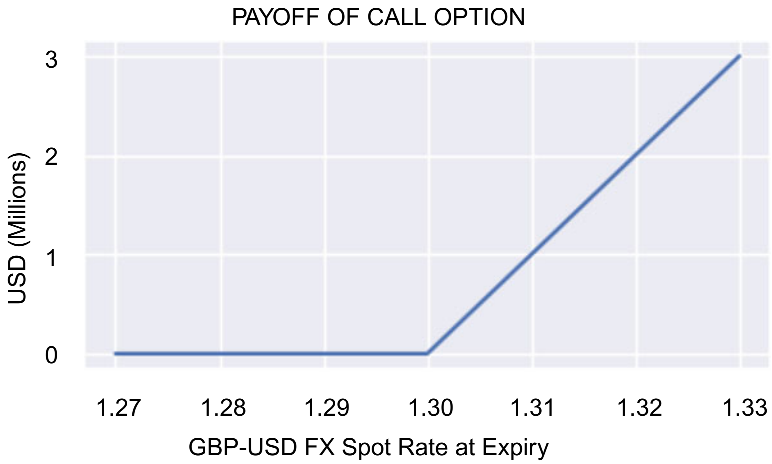
4.1 Option Payoff Basics

The payoff profiles of a vanilla call option and put option are shown in Fig. 4.1.

A call (put) option allows the trader to buy (sell) the underlying currency at a pre-specified price, called the strike price on the expiry date. It follows

---

[1] According to the Triennial Central Bank Survey on foreign exchange turnover in April 2019 by the Bank of International Settlements (BIS). The full report is available at https://www.bis.org/statistics/rpfx19_fx.pdf.



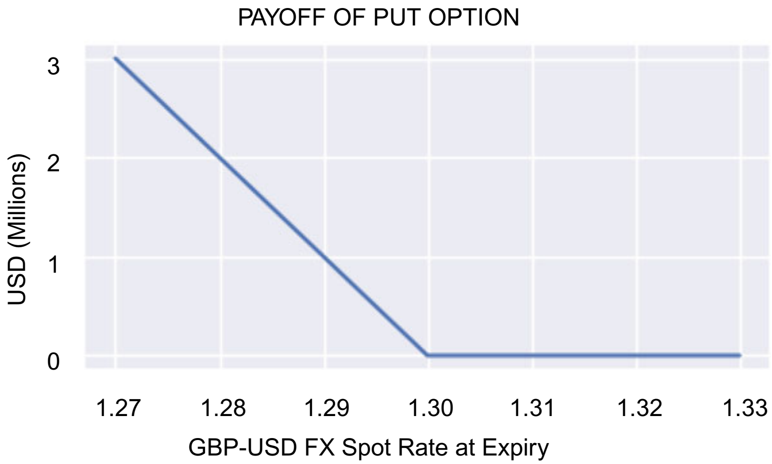


Fig. 4.1 The upper and lower figures show the payoff profiles of a call and put option respectively. In this example, the payoff corresponds to an option with a notional of 100 million GBP and a strike of 1.30. The horizontal axis shows the level of GBP-USD at the maturity, or expiry, time of the option. If GBP-USD is above (below) 1.30 on the expiry date of the option, the owner of the call (put) can buy (sell) 100 million GBP-USD at 1.30 if they so choose. The vertical axis shows the payoff to the owner of the option in USD. The payoff of the call option is zero if GBP-USD is below 1.30 at the expiry time because the trader would not rationally exercise her option to buy GBP-USD at the higher price of 1.30. However, it is positive if GBP-USD is above 1.30 and rises linearly. The put option behaves similarly, but the payoff is zero if GBP-USD is above 1.30, and positive if GBP-USD is below 1.30

that the payoff of the call option is given by

$$\max(S_T - K, 0) \times N,$$

where S_T is the price of the underlying security at the expiry of the option, K is the strike price and N is the notional, or number of units of the option purchased. Analogously, the put payoff is given by $\max(K - S_T, 0) \times N$.

To see this, suppose that the trader has purchased $N = 100$ million GBP of a $K = 1.30$ call option on GBP-USD. If $S_T = 1.31$, then the trader exercises her option and buys 100 million GBP at the strike price $K = 1.30$. She can then sell them in the market at $S_T = 1.31$ to receive $(1.31 - 1.30) \times$ 100 million USD $=$ 1 million USD. However, if $S_T < 1.30$ then she can abandon her option and receive nothing. The formula $\max(S_T - K, 0) \times N$ correctly describes her payoff.

The call payoff is large when the value of the underlying asset is high. The put payoff is large when the value of the underlying asset is low. However, losses are floored in that the payoffs of both calls and puts are never negative, regardless of the value of the underlying. Options therefore provide insurance against market moves and their buyer must pay a premium.

Purchasing a call (put) option is therefore a means to express the view that the underlying currency pair will rise (fall). However, since spot and forward contracts already provide investors with exposure to the direction in which a currency pair will move, this is unlikely to be the main use of options. Indeed, over the next two sections we will see that the raison d'etre of options is to provide investors with exposure to volatility risk.

4.2 Options as Bets on Volatility

To begin, I show that options are bets on volatility. If an investor purchases both the call option and the put option then her payoff profile is given by the V shape shown in Fig. 4.2. This strategy is called a *straddle*.

Unlike with calls and puts, it is clear that with a V shaped symmetrical payoff profile the investor is not concerned as to whether the underlying currency pair rises or falls because her payoff is the same if GBP-USD moves from 1.30 up to 1.31 or if it moves down to 1.29. In the example in Fig. 4.2 it is 1 million USD in both cases. Her main concern is that GBP-USD moves away from 1.30. The further it moves, whether up or down, the higher her payoff. If GBP-USD were to remain at 1.30, then her payoff is zero and she loses the premium she paid to purchase the options. More so than any other option portfolio, the straddle makes clear that options are really bets on

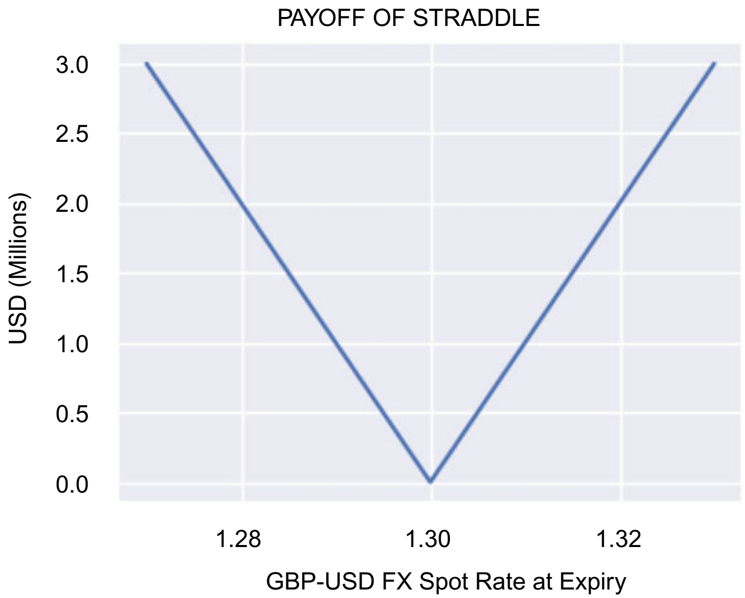


Fig. 4.2 The figure shows the payoff of a straddle option strategy in GBP-USD. This is the purchase of a call option and a put option both with the same strike and notional. I use a strike $K = 1.30$, and notional $N = 100$ million GBP each. The horizontal axis shows the level of GBP-USD at the expiry time of the option. The vertical axis shows the payoff to the owner of the option in USD

volatility, and not on direction because if GBP-USD volatility is high then the chances of GBP-USD moving away from 1.30 and thereby earning a higher payoff are higher. For this reason, the buyer of an option is said to be *long volatility* and the seller is said to be *short volatility*.

Next, I show that individual call and put options are also bets on volatility. Instead of the straddle, suppose that GBP-USD is trading at 1.30 and that the trader purchases a $K = 1.30$ GBP call option with $N = 200$ million GBP. The payoff of this position is shown in the left part of Fig. 4.3. Simultaneously, the trader sells 100 million GBP-USD spot at 1.30. The payoff of this transaction is shown in the right-hand diagram in Fig. 4.3. The selling of 100 million GBP-USD is called a *delta hedge*.[2] The payoff of the portfolio formed by purchasing the call option and executing the delta hedge is equivalent to that of a straddle, with the notional of the call and the put each equalling 100 million GBP. Accordingly, a delta hedged call option is a straddle, which we have already established above is a bet on volatility! An

[2] More precisely, the trader should sell 100 million of the GBP-USD forward contract to be *forward hedged*.

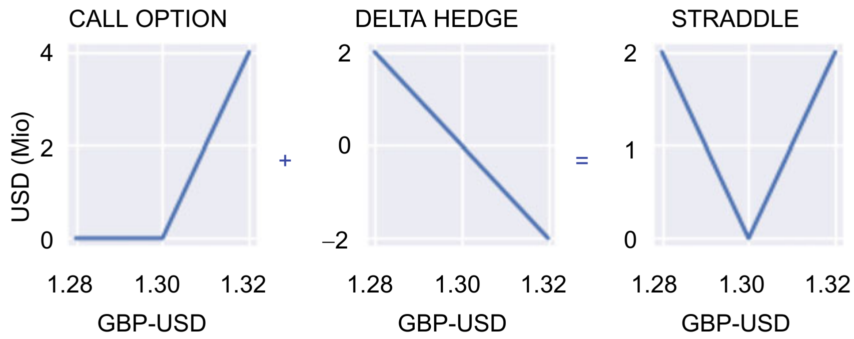


Fig. 4.3 The left figure shows the payoff of a call option on GBP-USD with strike $K = 1.30$ and notional $N = 200$ million GBP. The middle figure shows the payoff of selling 100 million GBP spot at 1.30. This is the delta hedge of the call option. The right figure shows that the payoff of a portfolio of the call option, plus delta hedge is identical to that of a straddle that is formed by purchasing $N = 100$ million GBP of a $K = 1.30$ put option, and $N = 100$ million GBP of a $K = 1.30$ call option

analogous result holds for the portfolio formed by purchasing the put option on GBP-USD and then delta hedging by purchasing GBP-USD spot.

The discussion here goes some way towards convincing the reader that, despite the common usage of call and put options to bet on the direction in which the underlying security will move, the ability to delta hedge means that options are fundamentally not bets on direction, but are bets on volatility.

4.3 Options as Volatility Hedges

Figure 4.4 recapitulates Example (iii) from Sect. 1.1.3. In the state where the BOE hikes the interest rate, and GBP-USD moves to 1.31, the total profit for the buyer of the option is 500 thousand USD. The reason is that the option payoff is 1 million USD, but the buyer paid 500 thousand to purchase it. Similarly, if the BOE did not hike the interest rate, and GBP-USD moves down to 1.29, then the total profit for the owner of the option is -500 thousand USD because the option payoff is zero, and the buyer paid $500 thousand USD to purchase it.

The key point to note is that the investor could have achieved exactly the same outcomes by trading spot instead of the option. If the investor simply purchased 50 million GBP in the spot market at 1.30, then her outcomes would be a gain of 500 thousand USD if GBP-USD moved to 1.31, and a loss of 500 thousand USD if GBP-USD moved down to 1.30. It is no surprise that this is exactly the delta hedge amount of the option that we

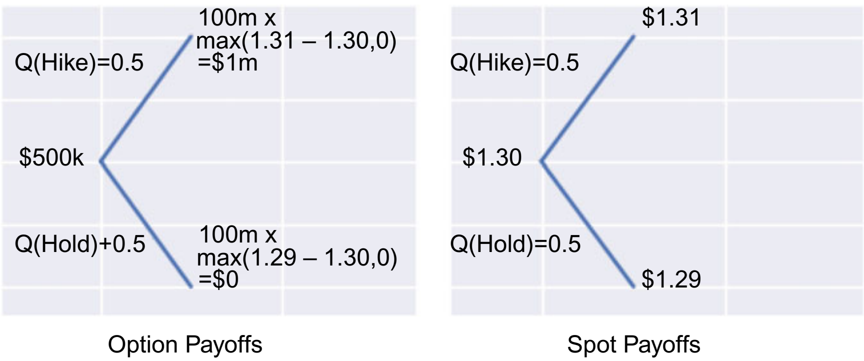


Fig. 4.4 The left plot shows payoffs of a 1.30 strike call option on GBP-USD with a notional of 100 million GBP. The right plot shows the payoffs of GBP-USD spot. Arbitrage based pricing implies that the price of the option is 500 thousand USD. This price exactly corresponds to expected value based pricing, but using the *Q* probabilities that implied from the price of GBP-USD spot

discussed in Sect. 1.1.3. This was the amount of GBP-USD that the seller of the call option purchased in the spot market to perfectly hedge herself. The option is said to be *replicable*.

The discussion up to this point leads to the following interesting question. Why should the option exist at all if its payoff is replicable? In this example, the investor may achieve the same profit and loss in each outcome by simply buying 50 million GBP-USD of spot. The answer lies in the assumption made that the terminal outcomes for GBP-USD are known to be 1.31 or 1.29. For example, suppose instead that we did not know ahead of time whether the outcomes would be 1.31 and 1.29, or a higher volatility set of outcomes, such as 1.32 in the case of a hike, and 1.28 in the case of no hike. That is, the volatility that the spot rate will exhibit is itself uncertain. Such a situation is known as *stochastic volatility*. In this case, it is no longer possible to perfectly replicate the option payoff using delta hedging alone. Put more precisely, there is no combination of delta and option premium that perfectly replicates the option payoffs.

To see this, consider Table 4.1. If the delta hedge amount is $x = +50$ million GBP and the option premium is $y = 500$ thousand USD then the payoff of the delta hedge is equal to the payoff of the option less the option premium if GBP-USD ends at either 1.31 or 1.29, as it did in Example (iii) in Sect. 1.1.3. However, if GBP-USD moves to 1.32, then the payoff of the delta hedge is +1 million USD, whereas the payoff of the option less the premium paid is 1.5 million USD. Similarly, if GBP-USD moves down to 1.28, then the payoff of the delta hedge is −1 million USD, whereas the

Table 4.1 The first column shows four states in which GBP-USD may expire. The remaining columns show the payoff of a delta hedge, the option payoff and the option premium paid in each of these states. The option is a 1.30 strike call on GBP-USD, and it is assumed that GBP-USD starts at 1.30

GBP-USD	Delta Hedge	Option Payoff	Option Premium
1.32	+2x	+2	−y
1.31	+1x	+1	−y
1.29	−1x	0	−y
1.28	−2x	0	−y

payoff of the option less the premium paid is −500 thousand USD. The reader may experiment with different values of x and y in Table 4.1, but will not find a pair such that the delta hedge and the payoff of the option, less the premium are equal at all values of GBP-USD. That is, adding the additional states of GBP-USD (or equivalently, adding stochastic volatility) means that the option is no longer replicable.

It should not be surprising that the option delivers better results than the delta hedge in the cases that GBP-USD moves to 1.32 or to 1.28. The reason is that the option was priced at 500 thousand USD assuming that the volatility of the GBP-USD spot rate would be contained to 1.31 or 1.29. Here, GBP-USD volatility has realised higher, and so the long option position benefits as expected.

It is clear that adding the possibility that GBP-USD may move to 1.32 or to 1.28 in response to the BOE decision should raise the market price of the option to a value higher than 500 thousand USD because if the value were 500 thousand USD, an investor may purchase the option and set $x = -50$ million GBP to implement an arbitrage. If GBP-USD expires at 1.29 or 1.30, her payoff is zero, but if spot expires at 1.28 or 1.32, her payoff is positive.

Suppose that 100 million GBP notional of the option is traded in the market at a price of 750 thousand USD. I return to how to get to a price of 750 thousand USD in the next subsection. If the investor buys the option at this price, and implements $x = -50$ million GBP-USD as her delta hedge, then her payoffs are as shown in Table 4.2. As discussed in Sect. 4.2, with the delta hedge, the payoff of the call option is symmetric. Finally, note that the profit is now a function of the outcome. If realised volatility performs well in that GBP-USD moves to 1.32 or 1.28, then the investor makes 250 thousand USD. However, if the volatility realised is low in that GBP-USD is contained to 1.31 or 1.29, then the investor loses 250 thousand USD. It is clear that the option is not a redundant security because such profit and loss outcomes are not replicable with a spot position alone. In real markets,

Table 4.2 The first column shows the four states in which GBP-USD will expire. The second column shows the payoffs of a short position in 50 million GBP-USD. The third column shows the payoffs of a long position in a notional of 100 million GBP of a 1.30 strike call option. The fourth column shows the option premium. Finally, the fifth column sums across the rows to show the profit from the strategy of purchasing the option and implementing the delta hedge. The units are in millions of USD. GBP-USD is assumed to start at 1.30

GBP-USD	Delta Hedge	Option Payoff	Option Premium	Profit
1.32	−1	+2	−0.75	+0.25
1.31	−0.5	+1	−0.75	−0.25
1.29	+0.5	0	−0.75	−0.25
1.28	+1	0	−0.75	+0.25

the fact that we do not know what volatility will be ahead of time and that options hedge volatility risk is reason for options to exist.

4.3.1 Dynamic Replication and the BSM Model

The discussion so far has focussed on a discrete set of outcomes in the FX spot rate. In the example presented, two possible outcome states represented the case where the volatility is known ex-ante, and four possible states represented stochastic volatility. The reader may wonder how this extends to real markets, where the range of FX spot outcomes more closely resembles a continuum of possible states.

The seminal work of BSM showed that the delta hedging approach to replicate the payoff of an option applies under a continuum of outcomes also, but only under the conditions that delta hedging occurs continuously in time and that the volatility of log-returns is constant across time. Constant volatility leads to a normal terminal PDF of log-returns, and therefore log-normal returns are required for the BSM formula to apply to option pricing.

The constant volatility assumption makes the BSM model analogous to our two state model. Indeed, Cox, Ross and Rubinstein (1979) developed the well known *binomial option pricing model*, which repeats the two state model over time to form a binomial tree with many possible outcomes. In the limit as each time step converges towards zero in size, so that there are infinitely many time steps in any finite time horizon, the terminal PDF converges to the log-normal, and the model therefore converges to the BSM model.

Upon first inspection, it may be confusing how the binomial model or BSM model is markedly different from the four state stochastic volatility model presented in this chapter. After all, the binomial model allows more than four final FX states, and the BSM model has infinitely many. Infinitely many states are more than four! How can it be that we can hedge a continuum of outcomes, but not hedge four? The key is to note that the binomial model of Cox, Ross and Rubinstein (1979) allows for delta hedging after every step. Therefore, the many final outcome states are only reached after many delta hedges are executed. The BSM model is simply the limiting case of this model, requiring delta hedging to occur continuously in time and therefore again many delta hedges are executed before reaching the final state.

The BSM model therefore has the striking implication that even though it is the inspiration for the multi-trillion dollar options business, if the model were true in a literal sense in that the volatility of spot log-returns were constant, then options are unlikely to exist at all, because they can be replicated via delta hedging. It is likely that algorithmic delta hedging *option replication* programs would offer option payoffs instead.

Despite the presence of stochastic volatility and therefore log-returns in FX markets that are often far from normally distributed, the BSM model remains central to FX options pricing and risk management. Next, I describe how options traders adapt the BSM model for practical trading.

Practical Adaptation of the BSM Model (Optional)

The remainder of this subsection assumes some prior exposure to the BSM model typically gained from a first course in option pricing. A full treatment of dynamic replication and the BSM model is beyond the scope of this chapter. However, I refer readers without such background to Iqbal (2018) to gain a fuller understanding of these ideas, and to Shreve (2000) for an understanding of the full mathematical machinery that underlies the derivation of the BSM model and stochastic volatility models. An understanding of the remainder of this subsection is not required to follow the remainder of the chapter.

First, to adapt the BSM model for option pricing, traders note that stochastic volatility leads to both skewness and kurtosis in the terminal PDF of the log spot return. Skewness refers to an asymmetry in the PDF of the spot FX price. Kurtosis or, more colloquially, fat tails, refer to there being more weight at both the centre, and in the tails of the PDF. This is shown in the lower part of Fig. 4.5. The figure compares the log-normal PDF with a *real* PDF. FX option prices must be adjusted relative to the BSM price to account for these features of the PDF.

In the PDF shown, the skewness of the distribution is to the left. Intuitively, this means that sell-offs in the FX spot rate tend to be larger and more volatile than when the FX spot rate rises. To adjust the BSM model to account for this, traders create a *volatility smile*. An example volatility smile is shown in the upper section of Fig. 4.5. The left part of the figure shows strikes below

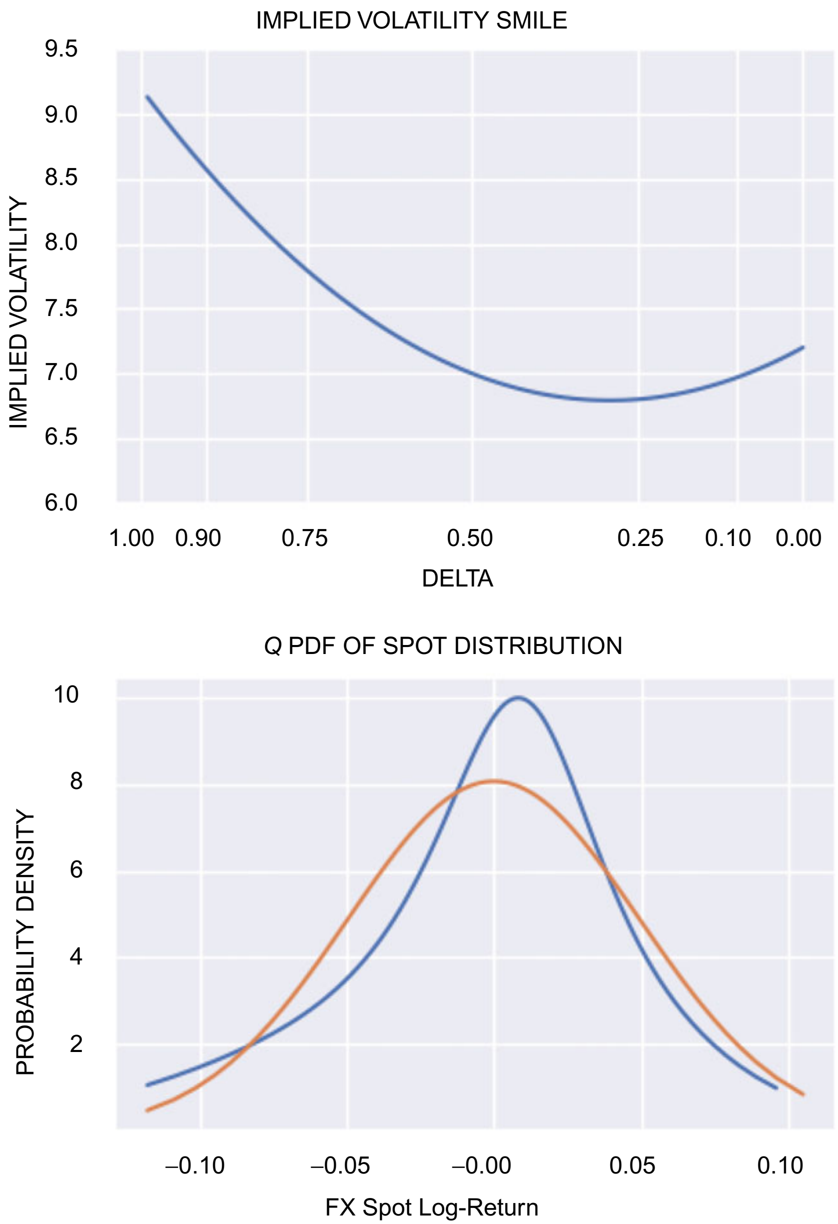


Fig. 4.5 The upper figure shows the implied volatility smile in delta space. Here, $\sigma_{\text{rr}} = -1\%$, $\sigma_{\text{atm}} = 7\%$ and $\sigma_{\text{bf}} = 0.3\%$. The lower figure shows how this smile maps into a *Q* PDF. I also compare this to the log-normal distribution associated with the BSM model. Note that there is a negative skewness compared with the log-normal

the forward and the right part of the figure shows strikes above the forward. In short, the implied volatility that must be inserted into the BSM formula to extract the price of low strikes is higher than that of the extract price of high strikes. This accounts for the volatility rising as the FX rate falls, or falling as the FX rate rises. Rebonato (1999) aptly wrote that BSM "implied volatility is the wrong number to put in the wrong formula to get the right price." The feature box later in the chapter studies the relationship between the volatility smile and the PDF in greater detail.

Second, to adapt the BSM model for risk management traders apply the so-called BSM *Greeks*. The Greeks are partial derivatives of the BSM option price with respect to parameters or variables in the model. For example, consider the Greek, *Vega*. It is defined as

$$\text{Vega} \equiv \frac{\partial C(\sigma_{\text{implied}})}{\sigma_{\text{implied}}},$$

where $C(\sigma_{\text{implied}})$ denotes the price of the option as a function of the level of implied volatility σ_{implied}. Within the BSM framework, it is odd to even consider this quantity because σ_{implied} is a (constant) parameter, and not a variable. However, traders understand that σ_{implied} does change on a daily basis, and wish to understand their profit and loss implications as it does. Vega is just one of the Greeks that FX options risk managers use and I refer readers to Iqbal (2018) for a complete analysis.

4.4 Extracting Probability Distributions from Options

Recall that in Example (iii) from Sect. 1.1.3 it was straight forward to extract the Q probabilities. The scenarios were that GBP-USD spot rate would move to 1.29 or to 1.31. Therefore, to satisfy Eq. (1.2) we required that $\text{Q(Hike)} = \text{Q(Hold)} = 0.5$, as shown in Eq. (1.4). In the presence of stochastic volatility, however, it is no longer possible to extract the Q probabilities from the spot market. Equation (1.2) becomes,

$$\begin{aligned} 1.30 = \text{Q}(1.32) \times 1.32 + \text{Q}(1.31) \times 1.31 + \text{Q}(1.29) \times 1.29 \\ + \text{Q}(1.28) \times 1.28. \end{aligned} \tag{4.1}$$

Here, $\text{Q}(S_T)$ refers to the risk-adjusted probability that GBP-USD expires at S_T. There are now too many unknowns to calculate the Q probabilities. For example, setting Q(1.32) = Q(1.31) = Q(1.29) = Q(1.28) = 0.25 satisfies Eq. (4.1), but so does Q(1.32) = Q(1.28) = 0.1 and Q(1.31) = Q(1.29) = 0.4. However, if the option price is known, then this price can help extract the Q probabilities of the spot rate.

For example, in the previous section, I assumed an option price of 750 thousand USD for a $N = 100$ million GBP-USD of the $K = 1.30$ call option. Applying Eq. (1.2) to the option gives us

$$0.75 = \mathrm{Q}(1.32) \times 2 + \mathrm{Q}(1.31) \times 1 + \mathrm{Q}(1.29) \times 0 + \mathrm{Q}(1.28) \times 0. \tag{4.2}$$

Here, the price of the option on the left-hand side and the payoffs on the right-hand side are quoted in millions of USD. Finally, making the assumption on the Q probability distribution that it is symmetric, meaning that $Q(1.31) = Q(1.29)$ and $Q(1.32) = Q(1.28)$ and imposing, of course, that the Q probabilities sum to 1, we find that $Q(1.32) = Q(1.31) = Q(1.29) = Q(1.28) = 0.25$ solves Eqs. (4.1) and (4.2). That is, knowing the traded price of the option has allowed us to calculate the Q probability distribution of spot, even in the stochastic volatility model.

The assumption that the Q distribution was symmetric was not necessary. However, without this assumption, the prices of options of different strikes are necessary to solve for the entire Q probability distribution.

This stylised example is intended to provide the reader with a framework to understand how option prices allow investors to uncover the Q probability distribution of FX spot. Subsection 4.4.2 translates these theoretical ideas into practical application.

4.4.1 The Volatility Risk Premium (VRP)

Recall the logic of Chapter 1, that assets with payoffs that provide a hedge against economic downturns earn a negative risk premium. Options are clearly such assets because volatility increases, often sharply and somewhat dramatically during economic crises. The owner of an option therefore expects to earn a negative risk premium. This risk premium may be expressed as a negative expected cash flow or expected return, as we have done over the previous chapters. However, in the case of options, it is often expressed as the difference between the standard deviation of the spot distribution that is priced into options, and that expected by investors. This is the VRP. The remainder of this subsection discusses this idea in the context of the example of the previous subsection, involving our one period GBP-USD option in the presence of stochastic volatility.

According to the theory of risk premiums, the expected payoff of the option under the objective P probability measure should be lower than 750 thousand USD. The objective probability of the larger moves, namely P(1.32) and P(1.28) should be lower than Q(1.32) and Q(1.31), and therefore the objective probability of smaller moves, P(1.31) and P(1.29) should be larger so that the probabilities sum to 1. Suppose, for example, that P(1.32) = P(1.28) = 0.1 and P(1.29) = P(1.31) = 0.4. Then the expected payoff of the option is

$$\mathrm{P}(1.32) \times 2 + \mathrm{P}(1.31) \times 1 + \mathrm{P}(\ 1.29) \times 0 + \mathrm{P}(1.28) \times 0 = 0.6.$$

As usual, the difference between the price of the option, and the expected payoff under the P probability measure is the risk premium. In this case, the owner of the option earns a risk premium of -150 thousand USD. The VRP is essentially this quantity, but expressed in terms of the standard deviation of the probability distribution of the FX spot rate. Let us discuss this idea.

Options investors often think of options in terms of the standard deviation that they imply for the FX spot rate. In our example, the standard deviation of the spot changes under the Q probability measure as

$$\begin{aligned}\mathbb{E}_t^Q\left[(S_T - S_t)^2\right]^{\frac{1}{2}} &= \left(\sum Q(S_T) \times (S_T - S_t)^2\right)^{\frac{1}{2}},\\ &= 1.58\%.\end{aligned}$$

However, under the objective P probability measure it is

$$\begin{aligned}\mathbb{E}_t^P\left[(S_T - S_t)^2\right]^{\frac{1}{2}} &= \left(\sum P(S_T) \times (S_T - S_t)^2\right)^{\frac{1}{2}}.\\ &= 1.26\%.\end{aligned}$$

Another way of expressing the risk premium is therefore to take the difference between the standard deviation derived from the market price of the option, 1.58%, and the objective standard deviation, 1.26%. It is clear that a larger difference between the P and Q standard deviations of the FX spot rate also corresponds to a larger difference between P and Q expected payoffs. Since *volatility* is the annualised standard deviation of the FX spot rate, as I discuss in the next subsection, when the above quantities are annualised, the difference is known as the VRP.

4.4.2 Practical Application with a Distributional Assumption

A quick and intuitive way to extract an approximate Q probability distribution of the future FX spot rate from FX options prices is to study so-called *implied volatility*. The concept of implied volatility originates from the BSM option pricing function. I explain this idea over the course of this section but as a heuristic, the reader may think of implied volatility as the annualised standard deviation of the forward looking Q probability density function (PDF) of the FX spot rate, at this stage.

For instance, suppose that an FX options market maker or other data provider quotes a cash price for a 1 year GBP-USD at-the-money forward (ATM) option.[3] The BSM function is perhaps most simply thought of as a mapping from implied volatility and the cash price. With the cash price in hand, the investor may apply the BSM function in reverse to back out the implied volatility.

Suppose that the cash price corresponds to an implied volatility of 8%. This means that, approximately at least, the FX options market implies that the FX spot rate in 1 year will be within ±8% relative to the 1 year forward price of GBP-USD, with a probability of 68%. The reason being that 68% of the probability distribution is contained within one standard deviation. Readers with less exposure in this area may find the refresher on probability distributions in Appendix B helpful.

The standard deviation of the PDF is proportional to the square root of time. For example, if the implied volatility of a 3 month option is 8%, then the FX options market implies that the FX spot rate in 3 months will be within $\pm\sqrt{3/12} \times 8\% = 4\%$ relative to the 3 month FX forward price. This square root relationship is derived as follows. First, note that the log-return over a period of n days can be split into daily log return as,

$$
\begin{aligned}
\ln \frac{S_{t+n}}{S_t} &= \ln\left(\frac{S_{t+1}}{S_t}\frac{S_{t+2}}{S_{t+1}}\cdots\frac{S_{t+n}}{S_{t+n-1}}\right)\\
&= \ln\left(\frac{S_{t+1}}{S_t}\right) + \ln\left(\frac{S_{t+2}}{S_{t+1}}\right) + \ldots\left(\frac{S_{t+n}}{S_{t+n-1}}\right).
\end{aligned}
$$

[3] ATM forward refers to an option, call or put, with strike equal to the forward at the expiry date of the option. That is, $K = f_t(T)$.

Next, take the variance on both sides of this equation.

$$\begin{aligned}\sigma^2\left(\ln \frac{S_{t+n}}{S_t}\right) &= \sigma^2\left(\ln\left(\frac{S_{t+1}}{S_t}\right)\right) + \sigma^2\left(\ln\left(\frac{S_{t+2}}{S_{t+1}}\right)\right)\\ &\quad + \ldots \sigma^2\left(\ln\left(\frac{S_{t+n}}{S_{t+n-1}}\right)\right)\\ &= n \times \sigma^2\left(\ln\left(\frac{S_{t+1}}{S_t}\right)\right),\end{aligned} \tag{4.3}$$

where $\sigma^2(x)$ denotes the variance of random variable x. In the first line I assume that daily returns are not autocorrelated, and in the second line, I assume that daily variance is constant. Finally, note that implied volatility, which I denote by σ_{implied} can be written as

$$\sigma^2_{\text{implied}} = 365 \times \sigma^2\left(\ln\left(\frac{S_{t+1}}{S_t}\right).\right) \tag{4.4}$$

The reason is that it is the annualised standard deviation of spot rate. Finally, substituting Eq. (4.4) into Eq. (4.3) we have,

$$\sigma\left(\ln \frac{S_{t+n}}{S_t}\right) = \sqrt{\frac{n}{365}}\sigma_{\text{implied}}, \tag{4.5}$$

which shows that the standard deviation of the log FX spot return scales with the square root of time, as required.

An important point to note is that the full (approximate) Q distribution was obtained using the price of only one single option, the ATM forward call or put option. This is analogous to the one period example in Sect. 4.4.

However, the reader will note my use of *approximate* in this subsection. There are at least two reasons. First, this reflects the fact that this approach assumes a log-normal PDF of the spot price. In real markets, PDFs are skewed and leptokurtic. To extract the full Q PDF of the FX rate, the prices of many options are required. I discuss this in the next subsection.

Second, the BSM model assumes that the volatility that will be realised is known ahead of time and that it is constant. Therefore, strictly speaking, the standard deviation of the Q distribution and P distribution in the BSM model is the same. That is, there is no VRP. In the next subsection, I provide a method that does not require the BSM formula or distributional assumptions to extract the Q probability distribution. However, the reader will find that practical implementation of the method described here is both convenient and adequate for many circumstances.

4.4.3 Practical Application without a Distributional Assumption

The *digital* or *binary* call is a derivative contract with payoff $\mathbf{1}_{S_T>K}$. Here, $\mathbf{1}_{S_T>K}$ denotes the indicator function, which takes the value 1 if $S_T > K$ and 0 otherwise. For example, if the investor were to purchase a digital call option on GBP-USD struck at 1.30 with a notional of 1 million USD, then her payoff is 1 million USD if the GBP-USD spot rate is at or above 1.31 at the expiry time, and zero otherwise (Fig. 4.6).

Digital contracts directly provide the Q distribution. To see this, apply Eq. (1.2) to the payoff. Letting $D_t(K)$ denote the price of the digital with strike K we have,

$$\begin{aligned} D_t(K) &= \mathbb{E}_t^Q\left[\mathbf{1}_{S_T>K}\right] \\ &= \mathrm{Q}(S_T > K) \times 1 + \mathrm{Q}(S_T < K) \times 0 \\ &= \mathrm{Q}(S_T > K). \end{aligned} \tag{4.6}$$

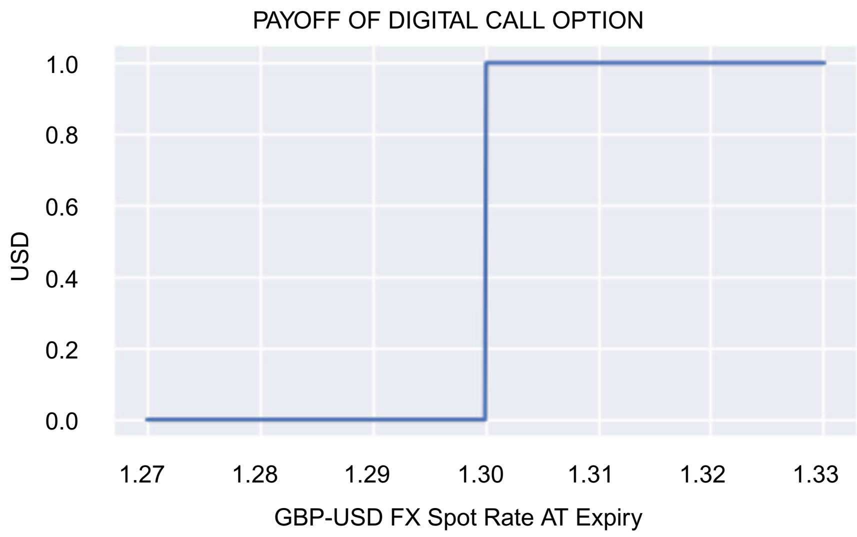


Fig. 4.6 The figure shows the payoff of a $K = 1.30$ GBP-USD call digital. The horizontal axis shows the level of GBP-USD at the expiry time of the option. The vertical axis shows the payoff to the owner of the option is 1 USD if GBP-USD is above 1.30, or zero otherwise

Therefore, by observing the traded price of digital contracts across the range of strikes, $D_t(K)$ for all K, the full cumulative distribution function (CDF) of the FX spot rate is obtained. Note that the price of many digital contracts, and not just one, is required to extract the Q distribution.

Later in this subsection, I show that a digital of strike K is replicated by buying a vanilla call option at strike $K - \varepsilon$, and selling a vanilla call option at strike $K + \varepsilon$, with notional $1/2\varepsilon$. This will lead to the relationship that

$$D_t(K) = -\frac{\partial C(K^*)}{\partial K^*}\bigg|_{K^*=K}. \tag{4.7}$$

Here, $C(K^*)$ denotes the price of a call option of strike K^*, ∂ denotes the partial derivative (see Appendix C.1 for a refresher on partial derivatives), and $|_{K^*=K}$ denotes that this partial derivative is evaluated at K.

Next, by applying Eq. (4.6), we have

$$\mathrm{Q}(S_T > K) = \frac{\partial C(K^*)}{\partial K^*}\bigg|_{K^*=K}.$$

Therefore, given the prices of the vanilla call options across strikes, the investor is able to extract the Q CDF. Finally, since the PDF is the derivative of the CDF (see Appendix B.1.2), the investor is able to extract the Q PDF also. That is

$$\mathrm{Q}(S_T = K) = \frac{\partial^2 C(K^*)}{\partial K^{*2}}\bigg|_{K^*=K},$$

where $Q(S_T = K)$ denotes the PDF. The feature box shows a practical implementation of extracting this Q distribution. However, the key point to note at this juncture is that just as in Sect. 4.4, if a distributional assumption is not made, the prices of *many* options are required to extract the Q probability distribution. Here, the prices of options at every strike K^* are required, and taking the second derivative with respect to K^* and evaluating this derivative at $K^* = K$ provides $\mathrm{Q}(S_T = K)$.

Digitals as Call Spreads

Consider the payoff profile that is generated by purchasing a call option of strike $K - \varepsilon$ and selling a call option of strike $K + \varepsilon$, as shown in Fig. 4.7. This commonly traded option strategy is called a *call spread*. In the example shown, $K = 1.30$ and $\varepsilon = 0.01$.

The call spread shown in Fig. 4.7 has a notional of 100 million GBP. Suppose instead that the notional is set to $1/2\varepsilon$. In this case the payoff would

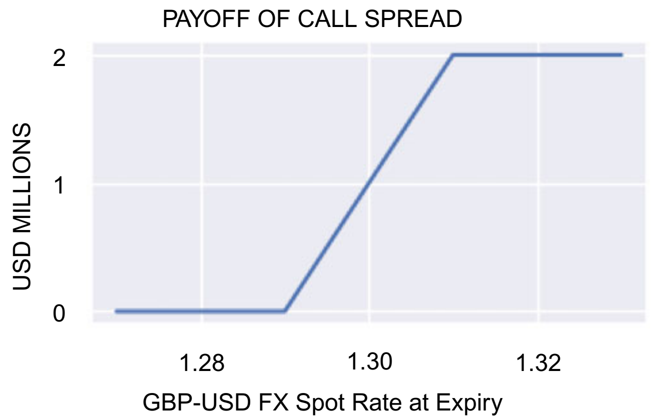


Fig. 4.7 The figure shows the payoff of a call spread. In this example, it is the purchase of a call option with strike 1.28 and sale of a call option with strike 1.32 each with notional of 100 million GBP. The horizontal axis shows the level of GBP-USD at the expiry time of the option. The vertical axis shows the payoff to the owner of the option in USD

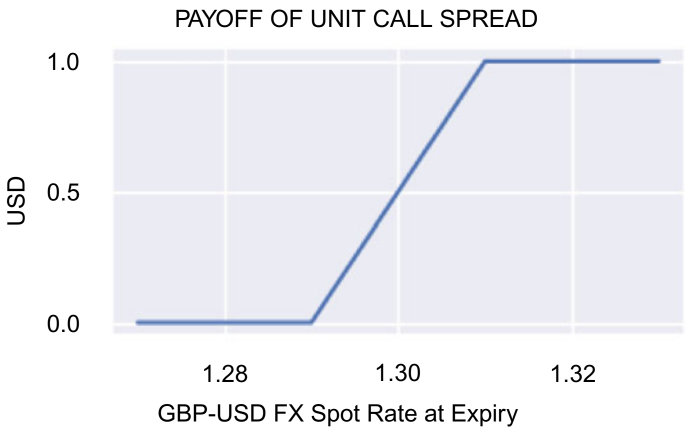


Fig. 4.8 The figure shows the payoff of a unit call spread. In the example, this is the purchase of call option with strike $1.30 - \varepsilon$ and sale of a call option with strike $1.30 + \varepsilon$ each with notional of $1/2\varepsilon$ EUR. Here, $\varepsilon = 0.01$. Like the digital, the maximum payoff is 1 USD

be as shown in Fig. 4.8. Already we see that the payoff of the unit call spread does not appear too dissimilar to that of the digital in that above 1.31 both the digital and the call spread pay 1 USD and below 1.29 they both pay nothing. The payoff is capped at 1 USD and hence I refer to this option strategy as a *unit* call spread.

The price of the unit call spread is therefore

$$\frac{1}{2\varepsilon}(C(K-\varepsilon, t, \sigma_i) - C(K+\varepsilon, t, \sigma_i)).$$

Next, take the limit as ε becomes infinitesimally small. As we decrease ε, the payoff converges to the profile shown in Fig. 4.6. In the limit as $\varepsilon \to 0$, the payoff of the unit call spread converges to that of the digital. The price of the digital payoff is therefore given by

$$D_t(K) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon}(C(K-\varepsilon) - C(K+\varepsilon)) = -\frac{\partial C(K^*)}{\partial K^*}, \tag{4.8}$$

which follows from the definition of a (calculus) derivative (see Appendix C.1). This is Eq. 4.7, as required.

Practical Extraction of the *Q* distribution (Optional)

This feature box details how the investor may extract the *Q* distribution from the prices of vanilla options. The discussion presented here is targeted at readers with some familiarity with the so-called *volatility smile* in FX options and the terminology relating to *delta space*. Readers without prior exposure in this area may consult Iqbal (2018) for a full treatment.

I denote the volatility of the at-the-money-forward (ATMF) option by σ_{atmf}, the 25 delta risk reversal by σ_{rr}, and the 25 delta butterfly by σ_{bf}. Finally, I impose a simple quadratic form for the volatility smile of

$$\sigma_{\text{implied}}(\delta) = \sigma_{\text{atm}} - 2\sigma_{\text{rr}}(\delta - 0.5) + 16\sigma_{\text{bf}}(\delta - 0.5)^2. \tag{4.9}$$

Here, δ refers to the delta of the call option. Note that with this functional form the following familiar equations hold at the $\delta = 0$, $\delta = 0.25$ and $\delta = 0.75$ points,

$$\sigma_{\text{atmf}} = \sigma_{\text{implied}}(50), \tag{4.10}$$

$$\sigma_{\text{rr}} = \sigma_{\text{implied}}(0.25) - \sigma_{\text{implied}}(0.75), \tag{4.11}$$

$$\sigma_{\text{bf}} = \frac{\sigma_{\text{implied}}(0.25) + \sigma_{\text{implied}}(0.75)}{2} - \sigma_{\text{atmf}}. \tag{4.12}$$

Equation (4.10) says that σ_{atmf} is the volatility of the 50 delta call. Equation (4.11) says that σ_{rr} is the implied volatility of the 25 delta call, less the volatility of the 75 delta call (equivalently the 25 delta put). Finally, Eq. (4.12) says that σ_{bf} is the average of the volatility of the 25 delta call and the 75d call (or 25d put), less the volatility of the ATMF option.

The upper chart in Fig. 4.5 plots Eq. (4.9), assuming that $\sigma_{\text{rr}} = -1.0\%$ and $\sigma_{\text{atm}} = 7\%$ and $\sigma_{\text{bf}} = 0.3\%$. This is the volatility smile. By applying the BSM formula to the volatility smile, I construct $C(K)$ over a range of K. Finally, I calculate the numerical second partial derivative to approximate Eq. (4.8) and extract the Q distribution of the spot rate. This is shown in the lower chart in Fig. 4.5.

The smile constructed in this example is a three point smile, so-called because three inputs, namely, σ_{atmf}, σ_{rr} and σ_{bf} were taken to be constructed. Most market making institutions in FX options use a five point smile, where the additional inputs are the 10 delta risk reversal points, $\sigma_{\text{implied}}(0.1) - \sigma_{\text{implied}}(0.9)$, and the 10 delta butterfly, $\frac{\sigma_{\text{implied}}(0.1)+\sigma_{\text{implied}}(0.9)}{2} - \sigma_{\text{atmf}}$. Some institutions also explicitly implement a no-arbitrage model such as the well known SABR model to calculate the functional form of $\sigma_{\text{implied}}(\delta)$, instead of that in Eq. (4.9). The investor may wish to carry out these tasks to fit the market with greater accuracy. However, Eq. (4.9) is sufficient for most practical investment decision-making, even if it is not for market making.

4.4.4 Correlation

Options markets in FX are unique among options markets in that complete triangles trade. By this I mean that one can trade an option on EUR-USD, an option GBP-USD, and also a EUR-GBP option. This is not true in other asset classes. For example, in the equity markets, one can trade an option on Microsoft-USD, an option on Apple-USD, but not on Apple-Microsoft (by which I mean Apple stock denominated in units of Microsoft Stock). The complete triangle makes it possible to extract FX *implied* correlation. The remainder of this subsection explains the meaning of this useful quantity.

The volatilities of EUR-USD, GBP-USD and EUR-GBP are related to each other through the correlation of the underlying currencies. To see this, first note that the EUR-USD spot rate is given by

$$S_t^{\text{EUR-USD}} = S_t^{\text{EUR-GBP}} \times S_t^{\text{GBP-USD}}. \tag{4.13}$$

This is an arbitrage relationship. The number of USD obtained from selling 1 EUR is $S_t^{\text{EUR-USD}}$. This must be the same as the number of USD obtained from first selling 1 EUR to obtain $S_t^{\text{EUR-GBP}}$ GBP, and then selling these GBP to obtain USD at $S_t^{\text{GBP-USD}}$.

Rearranging Eq. (4.13) to make $S_t^{\text{EUR-GBP}}$ the subject, and taking logs gives us,

$$\begin{aligned}
\ln \frac{S_{t+1}^{\text{EUR-GBP}}}{S_t^{\text{EUR-GBP}}} &= \ln\left(\frac{S_{t+1}^{\text{EUR-USD}}}{S_t^{\text{EUR-USD}}}\frac{S_t^{\text{GBP-USD}}}{S_{t+1}^{\text{GBP-USD}}}\right)\\
&= \ln \frac{S_{t+1}^{\text{EUR-USD}}}{S_t^{\text{EUR-USD}}} - \ln \frac{S_{t+1}^{\text{GBP-USD}}}{S_t^{\text{GBP-USD}}}.
\end{aligned}$$

Finally, take variances on both sides, and re-arranging the resulting equation gives us,

$$\begin{aligned}
&\rho\left(r_{t+1}^{\text{EUR-USD}}, r_{t+1}^{\text{GBP-USD}}\right)\\
&\quad = \sqrt{\frac{\sigma^2(r_{t+1}^{\text{EUR-GBP}}) - \sigma^2(r_{t+1}^{\text{EUR-USD}}) - \sigma^2(r_{t+1}^{\text{GBP-USD}})}{2\sigma(r_{t+1}^{\text{EUR-USD}})\sigma(r_{t+1}^{\text{GBP-USD}})}}.
\end{aligned} \tag{4.14}$$

Although not strictly correct, in practice traders often insert the ATM implied volatilities, $\sigma_{\text{implied}}(r^{\text{EUR-GBP}})$, $\sigma_{\text{implied}}(r^{\text{EUR-USD}})$ and $\sigma_{\text{implied}}(r^{\text{GBP-USD}})$ into Eq. (4.14) to back out a correlation, $\rho_{\text{implied}}\left(r^{\text{EUR-USD}}, r^{\text{GBP-USD}}\right)$ which is then referred to as the *implied correlation.*

FX investors may apply this formula to calculate implied correlations between currencies. Such calculations can serve as an input into covariance matrices in optimal portfolio construction problems, and in forming joint Q PDFs across currencies. In real FX markets there exists a large suite of so-called *correlation* products. The most well known are correlation swaps, dual and triple binaries, and worst-of options. A full treatment of these products is beyond the scope of this text. However, I mention them here to highlight their existence to the reader.

4.5 Real Exchange Rate Volatility

The previous chapters studied the main factors that drive the nominal FX rate, namely interest rates, inflation and the risk premium. However, motivated by the PPP model of FX, I assumed that the expected real FX rate remained stable. I did not necessarily assume that it was stable at 1.0, as strictly required by PPP but did assume that it remained fixed in that it exhibited no volatility. We did examine reasons for the real FX rate to deviate from 1.0 in Sect. 3.1.3, such as the Balassa-Samuelson theory. However, again, the motivation for this discussion was to understand persistent deviations in the real FX rate from 1.0, as opposed to volatility in the real FX rate itself. The purpose of this section is to study sources of volatility in the real FX rate.

I begin by again applying the central pricing equation, Eq. (1.10), to derive a relationship that will aid our thinking on real FX rate volatility. First, apply the central pricing equation from the perspective of investors in the foreign economy to assets in the foreign economy,

$$V_t = \mathbb{E}_t^P\left[\beta \frac{u'(c^f_{t+1})}{u'(c^f_t)} X_{t+1}\right]. \tag{4.15}$$

Here, c^f_t denotes the consumption of investors in the foreign economy. If, for example, the currency pair of interest is GBP-USD, then GBP is the foreign (FOR) currency and the USD is the domestic (DOM) currency. c^f therefore refers to the consumption of the U.K. investor, V_t is the price of an asset denominated in GBP, and X_{t+1} is its payoff denominated in GBP.

Next, consider the pricing of foreign assets from the perspective of domestic investors and apply the central pricing equation once more as,

$$\underbrace{S_t V_t}_{\text{Price in DOM}} = \mathbb{E}_t^P\left[\beta \frac{u'(c^d_{t+1})}{u'(c^d_t)} \underbrace{S_{t+1} X_{t+1}}_{\text{Payoff in DOM}}\right]. \tag{4.16}$$

The left-hand side in this equation is the price of the same foreign asset as in Eq. (4.15), multiplied by the FOR-DOM spot rate to give us its price in the DOM currency. Returning to the GBP-USD example, this is the number of USD that the investor must spend to purchase 1 unit of the asset.

The payoff in the DOM currency is its payoff in the FOR currency, multiplied by the FOR-DOM spot rate at time $t + 1$. This would be the number of USD the investor receives as a payoff at time $t + 1$.

Note that the utility function is now that of the domestic investor. An important assumption that enables us to write Eq. (4.16) is that there exist frictionless capital markets, allowing the domestic investor to allocate capital into foreign assets until this optimality equation is satisfied.

Here, I assume that foreign and domestic investors have identical utility functions. It may indeed be the case that preferences diverge across countries, and this may impact the real FX rate through inhomogeneous preferences, although this topic is beyond the scope of this chapter.

Finally, comparing Eq. (4.15) and Eq. (4.16) we have,[4]

$$\beta \frac{u'(c^f_{t+1})}{u'(c^f_t)} = \beta \frac{u'(c^d_{t+1})}{u'(c^d_t)} \frac{S_{t+1}}{S_t}, \tag{4.17}$$

which, after taking logs allows us to express the real FX log return in terms of foreign and domestic marginal utilities as,

$$\ln \frac{S_{t+1}}{S_t} = \ln \frac{u'(c^f_{t+1})}{u'(c^f_t)} - \ln \frac{u'(c^d_{t+1})}{u'(c^d_t)}. \tag{4.18}$$

The remainder of this section is devoted to building economic intuition around Eq. (4.18).

[4] The reader may note that in Eq. (4.17) I have simply equated the inside of the expectation operators in Eqs. (4.15) and (4.16). This is not correct in general. Strictly put,

$$\beta \frac{u'(c^f_{t+1})}{u'(c^f_t)} + \varepsilon^f_{t+1} = \beta \frac{u'(c^d_{t+1})}{u'(c^d_t)} \frac{S_{t+1}}{S_t} + \varepsilon^d_{t+1},$$

where ε^f_{t+1} and ε^d_{t+1} are random variables that are orthogonal to the available payoffs. That is, $\mathbb{E}^P_t[\varepsilon^f_{t+1} X_{t+1}] = 0$ and $\mathbb{E}^P_t[\varepsilon^d_{t+1} S_{t+1} X_{t+1}] = 0$. Readers with prior exposure to asset pricing theory will note that $\varepsilon^f_{t+1} = \varepsilon^d_{t+1} = 0$ in complete markets because the discount factor is unique and therefore Eq. (4.17) applies as it is written. Further, in an incomplete market, the projections of marginal utility growths on the space of available payoffs are equal across investors. The intuition here is that investors with identical utility functions should share as much risk as possible using the available assets. A treatment of this topic is beyond the scope of this book. I refer readers without prior exposure in this area to Cochrane (2005) for a full treatment. However, the intuition that I aim to convey in this chapter can be fully understood by taking Eq. (4.17) as given.

4.5.1 Fixed Real Exchange Rates and International Risk Sharing

If the real exchange rate is fixed, as it is in the PPP model, then $\ln \frac{S_{t+1}}{S_t} = 0$. Equation (4.18) then simplifies to

$$\frac{u'(c^f_{t+1})}{u'(c^f_t)} = \frac{u'(c^d_{t+1})}{u'(c^d_t)}. \tag{4.19}$$

This means that the marginal utilities of foreign and domestic investors move in lock step. If we assume that foreign and domestic investors have utility functions of CRRA form (see Eq. [1.29]), then Eq. (4.19) simplifies further to

$$\frac{c^f_{t+1}}{c^f_t} = \frac{c^d_{t+1}}{c^d_t}. \tag{4.20}$$

That is, foreign and domestic investors perfectly share macroeconomic risk. If foreign investors suffer a 10% drop in their consumption due to an adverse economic shock, then so do domestic investors, whether that shock occurs in the foreign country or in the domestic country.

This remarkable result is also intuitive, particularly in the context of the PPP model. Remember, we have assumed frictionless capital markets, which allow domestic investors to buy and sell foreign assets just as they would domestic assets. This was the motivation for Eq. (4.16). Further, recall that it is the ability to frictionlessly transport real goods internationally, as described in Sect. 3.1.2, that motivates the PPP model. If both goods and capital are frictionlessly mobile across international borders, and foreign and domestic utility functions are also identical, then there is no difference between being a foreign or a domestic investor. They are economically one and the same, and there is no reason that their investment portfolios, realised real returns and therefore consumption growths should diverge.

4.5.2 Volatility of Real Exchange Rates and Imperfect International Risk Sharing

The previous section argued that zero volatility in the real FX rate implies perfect international risk sharing across foreign and domestic investors. The reverse of this argument is that imperfect macroeconomic risk sharing corresponds to volatility in the real exchange rate. Brandt, Cochrane and Santa-Clara (2006) put it another way, expressing this idea as "real exchange rates move and their fluctuations blunt risk sharing." This concept may be familiar and intuitive to FX volatility traders because FX volatility typically rises when international economic divergences emerge.

The frictionless transport of real goods was an important assumption to establish perfect international risk sharing in the previous subsection. To begin the discussion, I show that frictions in transporting goods across borders must lead to imperfect international risk sharing.

Consider, a hypothetical (limit case) example involving the U.S. and China. Suppose that a tariff war escalated to such an extent that trade between the two became impossible. If a positive productivity shock then arrived in China through, for example, the advent of a new technology or another source, the Chinese equity markets would rise. However, U.S.-based owners of Chinese equities cannot benefit. Ultimately, if widgets are not transportable then all U.S. produced widgets must be consumed by U.S. based investors, and China produced widgets must be consumed by China-based investors. It is clear, therefore, that the macroeconomic risk of a Chinese productivity shock has not been equally shared. Equation (4.18) tells us that whenever international macroeconomic shocks are not shared, we experience real exchange rate volatility, because $c^f_{t+1} \neq c^d_{t+1}$ and therefore $\ln \frac{S_{t+1}}{S_t} \neq 0$.

To understand the economic intuition, I move from this hypothetical example to one that more closely resembles real economies. As discussed in Chapter 2, real economies produce a mixture of tradeable and non-tradeable goods. Suppose that China incurred a uniform productivity shock that boosted productivity across all sectors. In the frictionless model, there would be an outflow of goods from the China to the U.S. based owners of Chinese assets. However, due to the friction that not all goods are tradeable, some of the extra goods produced cannot flow out of China. The extra supply of goods in China causes their prices to fall. Therefore, China becomes cheaper relative to the U.S. and the real USD-CNH exchange rate strengthens.

Let us match this description to Eq. (4.18). In the case of USD-CNH, USD is the FOR currency and CNH is the DOM currency. The goods that cannot be transported from China to the U.S. must be consumed by China-based investors. Therefore, c^d_{t+1} must rise. Applying a concave utility function, as described in Sect. (1.2.1), leads to a fall in $\ln u'(c^d_{t+1})$ and therefore a rise in $\ln \frac{S_{t+1}}{S_t}$. That is, the real CNH weakens against USD, as required.

The Central Bank and Price Stability

Suppose that the central bank acts to maintain stable prices. In this example, the People's Bank of China (PBOC) would increase the money supply and/or cut the nominal interest to boost the falling prices of Chinese goods. If this were to happen, then the only way for the Chinese real FX rate to weaken is if the nominal USD-CNH FX rate were to rise.

China's REER

The reader may recall Fig. 3.2 that showed a steady climb in the Chinese Yuan's REER as the country's productivity has risen, over the course of the past 25 years. This data may seem to counter the theory presented in this section. Here, we found that an increase in productivity leads to a weakening real exchange range. How can this be consistent? There are at least two important points to note.

First, recall that we turned to the Balassa-Samuelson theory to explain Fig. 3.2. The Balassa-Samuelson argument is based on a differential in the productivity between tradeable and non-tradeable goods. It argues that increasing productivity in the tradeable goods sector relative to the non-tradeable goods sectors leads to a stronger real exchange rate. In essence, the example presented here assumes an increase in the productivity in the non-tradeable goods sector. The Balassa-Samuelson theory therefore predicts that the real CNH weakens against the USD. Although the Balassa-Samuelson theory and the one presented here are unrelated, they make consistent predictions on the impact on the real FX rate.

Next, one may ask how the approach presented here attempts to explain the data in Fig. 3.2. In the example, the real CNH weakened because of frictions in trade. Over the period shown, trade has become cheaper and grown significantly. The decrease in trade frictions over time could have led to a strengthening of real CNH, as productivity has increased.

Summary

The takeaway is therefore that FX investors must not assume that productivity increases are beneficial for the FX rate. They may indeed weaken the currency if those increases cannot be exported, as predicted by both the Balassa-Samuelson theory, and the model presented here. This is intuitive. After all, imagine if the opposite were true that the FX rate strengthened in response to a productivity increase in non-tradeable goods. For this to happen, foreign investors would be required to pay a higher price for the currency of a country whose productively produced goods they cannot consume.

Understanding Equity Market and FX Relative Volatility

Figure 4.9 compares realised volatility in the S&P 500 with that of the AUD-USD since the start of 2007. As the reader may expect, volatility rose in both the equity and FX markets during the GFC of 2008, and also during the Coronavirus crisis in 2020. These periods are highlighted by the vertical grey bars.

It is striking that during the Coronavirus crisis, the ratio of S&P 500 volatility to AUD-USD volatility rises substantially, to a factor of just under 4. However, the divergence in S&P 500 and AUD-USD volatility during the GFC was not as substantial, with the ratio rising to approximately 2. Let us interpret this data within the framework provided by Eq. (4.18) and our understanding of nominal FX rates built during the previous chapter.

If the consumption path of Australian and U.S. investors did not diverge during the Coronavirus crisis, then according to Eq. (4.18), there is no reason for the volatility of the real AUD-USD FX rate to rise. Since both economies locked down and re-opened broadly simultaneously, it is reasonable to suspect the consumption paths of U.S. and Australian investors remained well correlated through the period.

Note also that the broadly simultaneous closing down of the U.S. and Australian economies leads to lower inflation expectations in both economies, and dovish responses from the respective central banks. This led to less need for the nominal FX rates to adjust through the mechanisms described in the previous chapter.

Equity markets, however, adjusted lower as expectations for future corporate earnings fell, and then adjusted higher as expectations rose again as economies recovered. This led to high realised volatility and thereby a high ratio of S&P 500 to AUD-USD realised volatility.

The smaller rise in the ratio of equity market and FX volatility during the GFC suggests that the consumption paths of U.S. and Australian investors were less correlated than during the Coronavirus crisis. This is intuitive. The GFC can be crudely thought as a crisis that initially impacted the U.S. mortgage sector, before becoming a global banking crisis, and ending as a European periphery sovereign debt crisis. While the U.S. and Australian economies were impacted by the pandemic in a similar way, there is no reason to believe that their exposures to U.S. mortgages, the banking sector, or European periphery sovereign debt is similar. This meant that the consumption paths of Australian and U.S. investors may indeed have diverged during the GFC, and this led to higher real exchange rate volatility. AUD-USD realised volatility therefore kept pace with S&P 500 volatility in 2008.

4.6 Chapter Summary

- Options provide investors with exposure to volatility risk.
- If the volatility that will be exhibited by the spot rate is known ahead of time, then options are redundant securities, because their payoffs can be replicated through delta hedging. Options' raison d'etre is therefore that they complete the market for volatility risk.
- The existence of FX options themselves therefore implies that the BSM model cannot correctly describe how to price options. Traders therefore *adapt* the model for the purposes of option pricing by using a volatility smile, and for risk management purposes by applying the Greeks.
- Implied volatility is, to a good approximation, the annualised standard deviation of the Q PDF.
- We can extract the Q PDF using data from FX options prices. By making distributional assumptions, few or even just one option price is required to do this. However, to extract a full, model free Q PDF, data on the prices of options across strikes must be known.
- The performance of FX options is typically negatively correlated with assumptions. Option prices are therefore higher than their expected payoffs. This is known as the VRP.
- Real FX rate volatility is related to the degree of international risk sharing. The greater that the consumption paths of foreign and domestic investors diverge, the greater volatility one may expect in the real FX rate.
- The real FX rate may strengthen as productivity in the tradeable goods sector increases. However, it may weaken if the productivity increases occur in the non-tradeable goods sector.

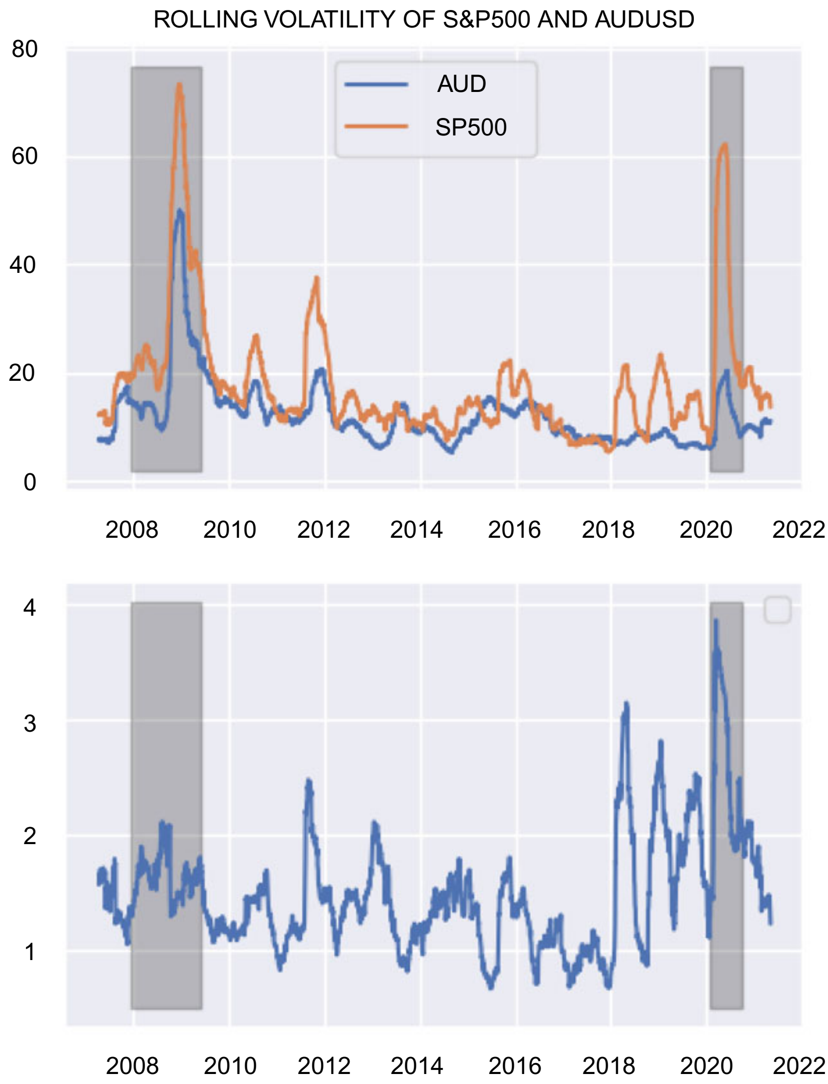


Fig. 4.9 The upper chart shows the 3 month rolling realised volatility of the AUD-USD FX rate and the S&P 500 index. The lower chart shows the ratio of these two series. The shaded areas indicate the GFC recession and the Coronavirus crisis respectively. The key observation is that the ratio of equity market volatility to FX volatility was lower during the GFC than during the Coronavirus crisis. Based on the theory developed in this chapter, one may suspect that Coronavirus risks were globally shared and in a more co-ordinated manner than those associated with the GFC

References

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637.

Brandt, M., Cochrane, J. H., & Santa-Clara, P. (2006). International risk sharing is better than you think, or exchange rates are too smooth. *Journal of Monetary Economics, 53*(4), 671–698.

Cochrane, J. H. (2005). *Asset Pricing* (Rev. ed.). Princeton University Press.

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics, 7*, 229–264.

Iqbal, A. (2018). *Volatility*. Wiley Finance Series: Practical Options Theory.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science, 4*(1), 141.

Rebonato, R. (1999). *Volatility and correlation*. Wiley.

Shreve, S. E. (2000). *Stochastic calculus for finance II: Continuous-time models*. Springer.

Part II

Macroeconomic Variables and Monetary Economics

Part I can be briefly recapitulated as follows. To determine the spot FX rate, the investor should (i) estimate the expected future FX rate, (ii) add in a risk premium to calculate the forward, and (iii) then use the CIP relationship to discount it back to the present time to calculate the spot FX rate. I argued that changes in the risk premium are the biggest driver of day-to-day FX spot price movements, that imperfect international risk sharing drives volatility in the real FX rate, and I described how changes in expected inflation impact (i) and (ii), and how changes in nominal interest rates impact (iii).

Part II studies the macroeconomic accounts and models of fiscal and monetary policy that influence the real exchange rate, the nominal interest rates that enter the CIP relationship, and inflation rates. In short, it relates the important variables that are required for FX determination in Part I to the macroeconomy.

To begin, Chapter 5 studies the balance of payments (BOP) accounts. In short the BOP says that countries that net export goods and services receive financial claims against the countries that are net consumers of goods and services. These claims are denominated in the currency of the debtor country. The reader will understand how the current and financial accounts of a nation are two sides of the same coin. I discuss how the experience of the U.S. shows that persistent current account deficits do not necessarily correspond to a weakening currency as one may intuit, so long as there is demand for the country's financial assets. This chapter also explores the basics of the government accounts and introduces the national accounts identity. This allows us to understand concepts such as *twin deficits*, and how they are related to the current account.

How should the BOP and national accounts impact the FX rate? To answer this question Chapter 6 introduces the extended version of the classical and seminal Mundell-Fleming model. This model also provides a simple framework in which to understand how monetary and fiscal policy may interact with the real FX rate. In short, the model extends the classical Keynesian IS-LM framework to incorporate the BOP identity. Although this approach suffers from some of the shortcomings of the IS-LM approach, which I discuss further in Chapter 7), it provides a framework to think about the interaction of macroeconomic accounts and the FX rate. It can also be used to derive the important Mundell-Fleming trilemma, which remains central in policymaker thinking, as well as in institution design such as that of the Euro area. Finally, the model is further built upon to allow for a changing price level in the discussion of inflation that follows in Chapter 7.

While Chapter 3 discussed the importance of expected inflation in FX determination, it did not provide any theories of inflation itself. This topic is broad enough to fill several textbooks, and therefore I do not provide an in depth study. However, in Chapter 7 the reader will gain sufficient understanding of the ideas embedded in the key theories, namely the IS-LM approach, monetarism's $MV = PY$ equation, New Keynesian Models and the Fiscal Theory of the Price Level, as well as concepts such as inflation measures, Taylor rules, and the Philips Curve that are frequently referenced in research reports and on trading floors.

5

Macroeconomic Aggregates and the FX Rate

Practitioners in FX markets are well versed with changes in macroeconomic aggregates as a narrative for FX price movement. For instance, an improving or worsening current account may be cited in industry and media reports as a reason for an FX revaluation. Alternatively, a change in fiscal policy or trade policy may be a source of an FX price change.

The reader may question how these narratives tie together with the drivers for FX price movements described in the previous chapters. Why should macroeconomic aggregates matter for the FX rate? In short, the reason is that while the variables of inflation, interest rates, and the risk premium remain the drivers of FX price movement as we have discussed, macroeconomic aggregates such as the current account and government sector accounts may signal something about how these variables may evolve in the future, and these signals therefore impact the FX rate.

The main macroeconomic aggregates are the balance of payment accounts, consumption, investment and government expenditure. Given its traditional association with the FX rate, this chapter spends the majority of its focus on the balance of payments (BOP) accounts. I explain how and why the current and capital accounts exactly equal the financial liabilities accrued by a country over a given period of time. I also discuss how a country's net international investment position is related to its balance of payments. The reader will become fluent in understanding these commonly referenced accounts.

However, I also argue that the balance of payments must be carefully interpreted before using it to make predictions about the FX rate. A current

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_5

account deficit need not imply that the FX rate is too strong. Since the current and financial accounts turn out to be two sides of the same coin, there is no need for the currency of a current account deficit country to weaken as long as there is demand for that country's financial assets.

The above implies that the current account does not causally impact the FX rate. To see this, note that while the U.S. is able to run persistent current account deficits without a negative or even notable impact on the U.S. dollar, it is not possible for many emerging market countries to do so without devaluation in their currency. The reason is that the countries that run a current account surplus do not anticipate an explicit default or a default via high inflation in the U.S. on the foreign obligations that it has accrued through running a current account deficit. They believe that U.S. deficits will therefore be repaid in the future, and are happy to continue to finance them in exchange for the investment opportunities that the U.S. provides.

The demand for U.S. financial assets has remained strong. However, elsewhere, and in emerging economies in particular, a worsening current account deficit may signal high future inflation, and this leads to currency devaluation through the mechanisms described in Chapter 3.

Next, I introduce the national income and product accounts (NIPA) identity. The NIPA identity breaks down GDP into consumption, investment, government expenditure and net exports. I use this identity to study commonly cited relationships such as *twin deficits* and their relationship with the current account. However, this identity will also prove central in linking together the FX rate with the macroeconomic aggregates via the extended Mundell-Fleming model in the next chapter.

This chapter also contains several important feature boxes that apply our knowledge of the BOP accounts to real world economic circumstances. I discuss how the BOP relates to former Federal Reserve Chairman Ben Bernanke's famous *global savings glut* hypothesis as a cause for the GFC. I also discuss its relationship with the famous Triffin dilemma that is often, and perhaps incorrectly, cited as the cause of the collapse of the post World War Two fixed exchange rate system, Bretton Woods.

Finally, I provide a short tour of key fiscal statistics, such as the net worth of the government, net operating balance, overall balance, and their primary versions. The reader will understand how changes in fiscal policy may impact these accounts, and the potential impacts of common policies such as the central bank's financing of the government, and QE. The chapters that follow will then allow the reader to translate these changes on the FX rate.

5.1 Balance of Payments (BOP)

The BOP accounts record the economic transactions that a country makes with the rest of the world during a period of time. Before diving into the detail, let us study a simple example of international trade that highlights two of its most important components, namely the current account, and the financial account and their interrelationship.[1] The key intuition is that the current account logs net exports, such as transactions in goods and services, and the financial account logs capital flows, such as transactions in stocks and bonds, and that the current account and financial account are approximately equal to each other. Consider the example in the next subsection.

5.1.1 Example: The Current Account and the Financial Account

Suppose that the U.K. is the domestic economy trading with Japan. A Japanese car manufacturer sells a car to a U.K. customer at a price of 15,000 GBP. What happens to net exports and net capital flow? It depends on what the Japanese car manufacturer does with the 15,000 GBP. Consider the following three scenarios.

Scenario 1
The Japanese manufacturer can keep the cash in a U.K. bank account, or invest it into GBP-denominated securities, such as U.K. government bonds, or shares in a U.K. public company. In this case, U.K. net exports have decreased by 15,000 GBP and, simultaneously, the net amount of capital lent by the U.K. abroad has decreased by 15,000 GBP. The import of the car is recorded in the current account as a debit of 15,000 GBP, and the increase in liabilities of the U.K to Japan is recorded in the financial account as a an incurrence of liabilities of 15,000 GBP. Importantly, the current account equals the financial account.

Scenario 2
Alternatively, the Japanese car manufacturer can buy 15,000 GBP worth of U.K.-manufactured goods. In this case, U.K. net exports, and therefore the current account, remain unchanged. The U.K. capital outflow, and therefore also the financial account, also remains unchanged. A U.K. resident handed over 15,000 GBP and a U.K. resident received back the same amount. Again,

[1] Note that industry and academic literature sometimes refer to the financial account as the capital account, or use the terms interchangeably. This chapter applies the definition used by the IMF. I describe these in detail over the course of the chapter.

note that the current account equals the financial account, because they are both zero.

Scenario 3

Finally, the Japanese car manufacturer may exchange her 15,000 GBP for Japanese Yen or a third currency. However, this does not change the situation. The counterparty to which the Japanese manufacturer hands over the pounds must do something with those pounds. He can save the money in a GBP bank account, invest it into sterling-denominated securities or buy U.K.-manufactured goods. In the end, again, the current account equals the financial account.

To summarise, if a foreign resident sells goods to the U.K., then there are only three things that can happen to the money paid: foreign residents can buy British goods, invest in the U.K. or buy U.K government debt. In each case, U.K. net exports must equal U.K. net capital outflow and therefore the change in the current account must equal that in the financial account.[2] With this intuition in place, I discuss these accounts in more detail.

5.1.2 BOP Accounts

An economic transaction enters the BOP accounts when it takes place between a resident and a non-resident. Here, residency does not refer to citizenship, but to the location of the economic interest. Residents include individuals or households present in the nation for a period (typically more than one year), enterprises with significant production within the nation, and the government, among others.

The BOP is a *flow* statement in that it records transactions over a set period of time. The website of the International Monetary Fund (IMF) provides records as frequently as monthly.

A related account is The International Investment Position (IIP). It records the sum of the financial account of the BOP and valuation changes across time. It therefore records the stock of assets and liabilities that a nation has with the rest of the world at a moment in time. I return to the IIP in Sect. 5.3.

The BOP consists of the current, capital and financial accounts. The remainder of this section studies these three components.

---

[2] The full and correct BOP accounting identity is that the sum of the current account, capital account, financial account and errors and omissions equals zero. I discuss this over the course of the chapter.

5.1.3 The Current Account

The current account records transactions in goods and services, the primary income account, and the secondary income account. Goods and services are usually the largest components of the current account. The meaning of *goods and services* is relatively straight forward and so I do not further expand on it here.

Primary Income Account
The primary income account records the returns accrued to residents, and paid to non-residents from factors of production by which I loosely mean labour, financial assets and national resources. For example, the payments received when a resident works abroad enter as a credit, and the payments made to a non-resident working in the country enter as a debit. Further, income from investments in foreign stocks, bonds, other investments and rents on foreign property are also reported in the primary income account.

Secondary Income Account
The secondary income account records current transfers: transfers of goods, services or cash. These include remittances that residents send to their family and other connected people residing abroad. Another item under secondary income is the taxes on income and wealth held abroad. Finally, the secondary income account includes social benefits and social contributions such as donations of food and medicines, and donations to the government.

Current Account Balance
To summarise, the following equation defines the current account balance, *CAB*.

$$CAB = X - M + PIB + SIB, \tag{5.1}$$

where X denotes total exports, M denotes imports, PIB denotes the primary income balance, and SIB denotes the secondary income balance. The quantity $X - M$ is sometimes shortened to NX to denote net exports. The following feature boxes aim to provide the reader with some intuition around the CAB by studying it for several countries using data from the website of the IMF.

Current Account Example

The table below shows data from the website of the IMF[3] for the U.K. in 2020.

U.K.	Goods	Services	Total
X	399	343	742
M	548	205	753
PIB			–49
SIB			–36
Total	–149	137	–96

The U.K. net imported 149 billion USD of goods, but exported 127 billion USD of services. Therefore, while $X - M = -12$ billion USD, a relatively small quantity, adding on the negative *PIB* and *SIB* numbers mean that the U.K ran a sizeable deficit of $CAB = -96$ billion USD in 2020. With a Gross Domestic Product (GDP) of approximately 3 trillion USD, this amounts to a current account deficit of approximately 3% of GDP. In the next feature box, I describe how the U.K. has run persistent deficits and has therefore accumulated a large negative net IIP with the rest of the world. The cost of financing this is a major reason for the negative *PIB* it faces.

A helpful exercise is for the reader to familiarise themselves with the IMF database by matching off the data in this table with that in the Standard Presentation of the BOP, available on the website of the IMF.

Current Accounts in Practice

Figure 5.1 shows the CAB expressed as a percentage of GDP across time for five major economies, namely the U.S., the U.K., China, Japan and Germany. The important take away is that the U.K. and U.S. have largely run persistent current account deficits since the 1980s, whereas Germany, Japan and China have largely run surpluses. That is, China, Japan and Germany export more

[3] http://data.imf.org.

goods and services than they import, whereas the U.S. and U.K. do the opposite.

Why are the current account surplus nations willing to support the apparent profligacy of the deficit nations by exporting more goods and services, year after year, than they receive back? The reason is that the residents of nations running a current account (plus capital account) surplus build up financial assets in deficit nations. We already encountered such a situation in the the example involving the export of a car from Japan to the U.K in Sect. 5.1, and I discuss the relationship between the current and financial accounts in more detail over the course of the remainder of the chapter. However, the key point is that trade always balances in that the current account (plus capital account) must always equal the financial account. The residents of the current account surplus nations therefore accumulate stocks, bonds and other investments in the current account deficit nations. Since such financial assets are claims on future production, surplus nations therefore expect to net receive goods and services in the future. Put differently, the current account deficit nations have borrowed goods and services, and these need to be paid back in the future.

The long persistence of the current account surpluses and deficits may lead the reader to question whether they will ever reverse. However, the only *way out* for the U.K. and U.S. from the large negative net IIP that each has built up, without simply running a surplus in the future, is to either explicitly default or inflate away the value of the financial liabilities that they have against their creditors. As things stand, both countries have central banks and fiscal institutions that provide sufficient confidence to foreign residents to purchase their financial assets in exchange for goods and services and expect that they will be paid back in the future.

This confidence may not always remain, and it is certainly not the case everywhere. Turkey, for example, has allowed much higher and more volatile inflation. To finance Turkey's current account deficit, foreigners have therefore demanded much higher compensation, and this has been achieved through a higher nominal interest rate, and a devaluation in the Turkish Lira. The ability to run persistent current account deficits therefore rests, partly at least, on the credibility of the fiscal and monetary institutions in the country.

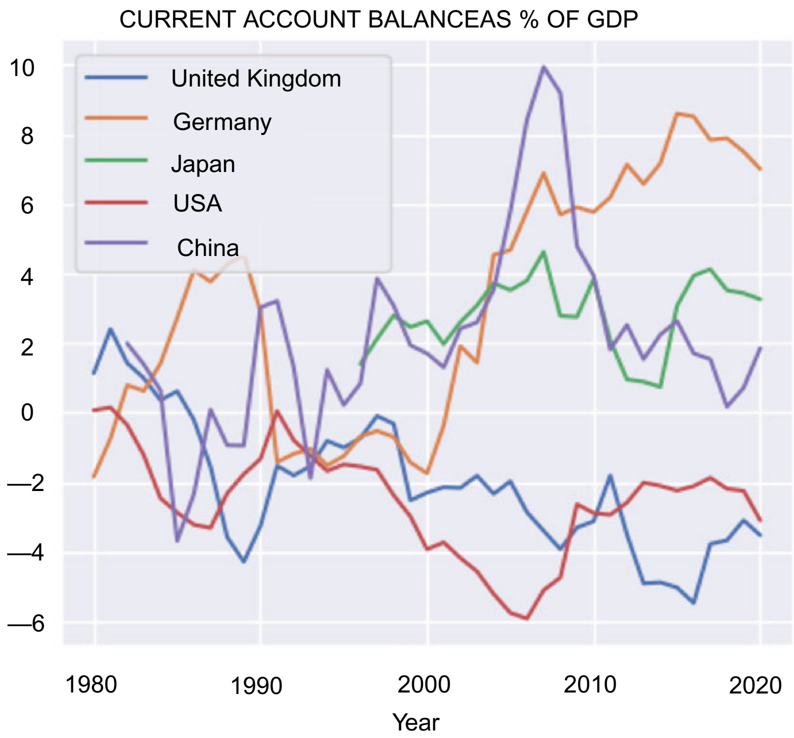


Fig. 5.1 The figure shows the current account balance expressed as a percentage of GDP in the U.K., Germany, Japan, U.S. and China since 1980. Germany, Japan and China have typically run a current account surplus, meaning that they export more in goods and services than they import. Meanwhile, the U.K. and U.S. have run persistent current account deficits (*Source* Data from website of the IMF)

5.1.4 The Capital Account

The capital account is typically much smaller than the current account. It records transactions that affect the stock of financial and non-financial assets and liabilities, but that do not originate from current account transactions. Suppose, for example, that a company buys the right to use a specific radio bandwidth from a foreign government, or buys property rights abroad, then there is no entry into the current account because these are non-produced assets. However, in an analogous way to an import of goods, there is an increase in financial liabilities to the foreign government.

Debt forgiveness, compensation that a country may make for, say, an act of war or an oil spill, and contributions towards international organisations such as NATO are all further examples of capital account transactions. Such payments are capital transfers that change financial liabilities. Again, since there are no produced goods or services involved, they would not naturally be recorded in the current account.

The sum of all the credit and debit entries in the capital account gives us the capital account balance, KAB. The sum of the current and capital account balances, $CAB + KAB$, provides a useful quantity; it is the amount that the nation must pay or receive from the rest of the world. If this quantity is positive, then the country is running a *surplus* in the sense that it is increasing its financial claims on foreign residence. The opposite is true if $CAB + KAB$ is negative.

A further point of note is that the above assumes that financial claims are exchanged for the $CAB + KAB$ surplus (deficit). If cash is received (paid), then foreigners claims on the surplus nation go down (up). In either case, such transactions are recorded in the financial account.

The next subsection discusses the financial account in more detail. However, at this stage, we may write down the accounting identity that relates the CAB, KAB and the financial account balance, denoted by FAB. It is,

$$CAB + KAB = FAB. \tag{5.2}$$

In the example of Sect. 5.1 there was only a current account entry, and no capital account entry because the example referenced a produced good. There, $CAB = FAB$. This section shows that KAB also needs to be considered.

Net Errors and Omissions

Finally, the reader should note that, although Eq. (5.2) is in theory an identity, it does not hold perfectly in practice. Net Errors and Omissions, NEO reflects the imbalances resulting from imperfections in source data and compilation of the balance of payments accounts. It is defined as

$$NEO \equiv FAB - (CAB + KAB).$$

NEO is usually small. Suppose that $NEO > 0$. This means that either the credit entries in the balance of payments (exports, other payments received, sale of assets, or incurrence of liabilities) were under-reported, or that debit entries (imports, other payments, or acquisition of assets, or the decrease in liabilities) were over-reported, or both.

5.1.5 The Financial Account

The financial account records financial transactions, classifying them into direct investment, portfolio investment, financial derivatives and employee stock options, other investment and reserve assets.

Unlike the capital and current accounts, where entries are recorded as credits and debits, the financial account records each transaction as the acquisition of financial assets or incurrence of liabilities. For instance, in the example of Sect. 5.1, if the Japanese exporter used the 15,000 GBP to purchase U.K. government bonds, then the U.K financial account would record an incurrence of liabilities, and the Japanese financial account would record an acquisition of financial assets. The financial account balance is,

$$FAB = \text{Net Acquistion of Financial Assets} - \text{Net Incurrence of Liabilities}.$$

Next, I describe each of the components of the financial account.

Direct Investment

Direct investment refers to a capital investment into an enterprise by a non-resident investor, where the investor exerts a significant degree of influence on the management of that enterprise. The investment may occur via equity or debt instruments. The reinvestment of earnings also falls under direct investment.

Portfolio Investment

Portfolio investment refers to all remaining cross-border transactions in debt or equity securities that are not classified as direct investment. For example, while an injection of capital towards a foreign subsidiary would be direct investment, the acquisition of foreign equity aimed at diversifying a portfolio of financial assets is classified as portfolio investment. Similarly, the purchase of foreign government bonds is a portfolio investment.

Financial Derivatives and Employee Stock Options, Other Investment, and Reserve Assets

Employee stock options allow employees to purchase stock in their company at a pre-specified price in the future. These are typically small amounts in the context of the BOP accounts.

Other investments include currencies and deposits, trade credits and loans. For example, the private acquisition of a foreign currency is not accounted for in direct investment, or portfolio investment because it does not involve a debt or equity instrument. Trade credits occur when a resident exporter gives to a non-resident importer, or vice versa. Loans are typically bank loans made to or by non-residents.

The last, and perhaps most important item for FX investors in the financial account is reserve assets. These are assets that are under the control of the central bank, and that are available for the central bank for direct financing of payment imbalances. These include monetary gold, special drawing rights, a reserve position with the International Monetary Fund, and other reserve assets.

Finally, note that all entries into a nation's BOP are recorded in a single currency. This could be the country's domestic currency, or one that is widely used by that country. For example, several Eastern European countries report their BOP in Euros, and many countries report in USD, given the USD's central role in payments.

5.2 Analytical Presentation of the BOP

The IMF provide both the *standard presentation* of the BOP that we have already discussed, and an *analytic presentation* of the BOP. The analytic presentation is designed to highlight items that may signal macroeconomic distress that arises from the external sector of the economy. It does so by moving three main items in the BOP to a separate *below the line* section called *Reserves and Related Items*. The items are reserve assets, net loans from the IMF, and exceptional financing.

This is best understood through an example. In the aftermath of the GFC, several distressed economies, including Greece, accepted loans from the IMF. A summary of the 2010 standard and analytic BOP of Greece are shown in Table 5.1.

The table shows that Greece ran a substantial current account deficit in 2010. The combined $CAB + KAB = -27.49$ billion USD. This was financed by financial account balance of 14.17 billion USD, plus an IMF loan of 13.74 billion USD. The reserve assets are -201 million USD, and these are therefore not enough to pay back this loan. It is clear from the analytic presentation that the current account deficit must be consolidated in the years to follow, although this may not be evident from the standard presentation. In the case of most countries, such a consolidation may occur through the channel of an FX devaluation so as to increase net exports. Since Greece and several other distressed economies were part of the Euro currency union, such a devaluation was not able to freely occur and this is may have been central to prolonging the Eurozone crisis (see, for example, Shambaugh [2012]).

Table 5.1 The table shows the standard presentation and the analytic presentation of the BOP for Greece in 2010 in millions of USD. The analytic presentation shows that the financial account balance that supports the current account deficit is not as strongly supported as one may expect based on the standard presentation, because 13.7 billion USD was obtained through IMF loans

	Standard presentation	Analytic presentation
Current Account	−30,262	−30,262
Capital Account	2,776	2776
Financial Account	−28,104	−14,168
Net Errors and Omissions	−617	−617
Reserves and Related items		−13,934
Reserve Assets		−201
Net IMF loans		13,735
Exceptional Financing		0

Source Website of the IMF

In this example, no *exceptional financing* was used. The reader may think of exceptional financing as comprising of debt forgiveness and intergovernmental grants, other borrowing for balance of payments support, debt restructuring and refinancing, debt prepayment and debt buyback, accumulation or repayment of arrears, and debt-for-development swaps. In summary, instruments that are intended to alleviate the burden of external debt.

5.3 International Investment Position (IIP)

The IIP accounts look similar to the financial accounts in their structure in that they are broken down into external assets and external liabilities, and then subdivided in the same way as the financial account into direct investment, portfolio investment, financial derivatives and employee stock options, other investments, and reserve assets. However, while the financial account reports flows during a period, the IIP reports the value of the stock of financial assets and liabilities of an economy to the rest of the world at a snapshot in time. This snapshot is typically taken at the end of a period, such as quarterly, or yearly.

The Net International Investment Position (NIIP) is the difference between external assets and external liabilities. The assets and liabilities are recorded at their value at a specific point in time. This important quantity summarises whether a country is a net lender or net borrower from the rest of the world at that point in time. The country is a net lender if NIIP is positive, and borrower if NIIP is negative.

5.3.1 Relationship Between NIIP and BOP

It is intuitive that the NIIP should be related to the BOP. If a country accrues financial assets during a period by, say, running a current account surplus, such that $FAB > 0$, then its NIIP should improve. The NIIP is therefore related to the cumulative financial account balance over time.

However, this is not the full picture because we must also include *valuation changes* that may occur over the period. For example, if a country's central bank holds a large stock of a foreign currency, and the value of that currency rises, then that country's NIIP rises commensurately. More generally, all of the financial assets and liabilities must be valued and their valuation changes included in the NIIP.

Finally, note that the NIIP also includes *other volume changes*. For example, if past loans issued by resident investors were to default, then this would not be included in the financial account balance, but would worsen the NIIP. This can be argued to be similar to the value of an asset going to zero.

Alternatively, if the central bank purchased foreign currency from a resident, then this transaction would not be recorded in the BOP, because it takes place between two residents. However, it would change some entries in the NIIP accounts, with the reserves of the central bank increasing, and those of residents decreasing.

The relationship between the NIIP and the BOP can therefore be summarised as

$$\Delta NIIP = FAB + VC + OVC, \tag{5.3}$$

where, $FAB = CAB + KAB + NEO$, VC denotes valuation changes, OVC denotes other volume changes and $\Delta NIIP$ denotes the change in the NIIP over the period.

5.4 Financing the BOP

The BOP can be considered to be financed if the overall balance in the analytical presentation of the BOP is not negative. That is, there is sufficient demand for a country's financial assets from private non-residents to cover a $CAB + KAB$ deficit.

The U.S. provides an example of a circumstance where financing of the BOP has remained strong over an extended period, and has actually strengthened during U.S. and global crisis periods. This has led to strength in the

USD and its status as safe-haven currency. I discuss this in detail in the following feature boxes.

For countries that have insufficient demand for their financial assets to cover current account deficits, something must change. The country may be able to increase the interest rate to increase demand for its assets, or it must contract the current account deficit. The most obvious mechanism to do so is through a devaluation in the currency. Alternatively, a combination of a weaker currency and higher interest rate may be required. It is for this circumstance that FX investors often study and monitor the BOP accounts.

BOP, IIP and the Global Savings Glut (GSG)

Much discourse around the large U.S. current account deficit focusses on competitiveness in goods and services. For instance, under the *America First* economic policy during the presidency of Donald Trump, the U.S. attempted to reduce its current account deficit and increase the competitiveness of its domestic industry by imposing tariffs on imported goods. It famously started with solar panels and washing machines in January 2018, and extended to aluminium, steel and other goods over the following months and years. China and other countries retaliated by imposing tariffs on U.S. exports. FX markets saw considerable volatility during 2019 as the rhetoric continued to escalate, most notably between the U.S. and China.

However, with the knowledge gained in this chapter so far, the reader will recognise that the current account and the financial account are two sides of the same coin. Therefore, to understand the U.S. current account deficit, rather than questioning why non-U.S. residents net sell goods and services to U.S. residents, it is equally valid to ask, why do non-U.S. residents wish to lend to U.S. residents? After all, since Eq. (5.2) must hold, a deficit in the current account is a surplus in the financial account!

This is the approach in reasoning that was taken by former Chair of the Federal Reserve, Ben Bernanke, in a famous speech in which he coined the now widely used term, *global savings glut*, to describe the increasing global savings from abroad that are held in U.S. assets (Bernanke 2005).

Bernanke cited several drivers of this desire for many parts in the industrialised and developing world to save, and the U.S. to be the recipient of these savings.

Among the industrial nations, such as Germany and Japan, Bernanke cites the ageing populations relative to the U.S. resulting in a greater desire for saving in those nations than in the U.S. This desire is not met domestically due to high ratios of capital to labour leading to fewer domestic investment opportunities and therefore to flows of capital into U.S. assets.

Among developing nations, Bernanke argues that financial crises such as Mexico in 1994, a number of East Asian countries in 1997–1998, Russia in 1998, Brazil in 1999 and Argentina in 2002 have encouraged emerging market central banks to build up FX reserves.[4] Indeed, Prasad (2015) makes a similar argument, writing that developing nations have come to view "international investors as fair-weather friends" who, "at the first sign of trouble....tend to turn tail, often precipitating asset market and currency crashes. Large stocks of foreign exchange reserves give emerging market policymakers some reassurance that they can better cope with capital flow and currency volatility".

If developing nations therefore wish to hold a "war-chest" of FX reserves, then where should they hold them? Bernanke argues that the U.S. is an attractive investment destination due to its productivity, the depth and sophistication of the country's financial markets, the special international status of the dollar in international payments, as well as its low political risk, strong property rights and developed regulatory environment.

Costs and Benefits of Imbalances

The benefits of imbalances are no different to any free market consumption-investment decision to buy or sell an asset. The saver prefers to save today and consume in the future, whereas the borrower prefers the opposite. Both parties may benefit, and the investment can increase productive capital in the economy over time; for example South Korea ran large trade deficits throughout the 1970s, it borrowed to invest, and it has become a success story for economic growth. In domestic markets, the interest rate equilibrates the price of this interaction, whereas in international imbalances it is both the interest rate and the FX rate.

[4] More recent crises, such as Argentina in 2020, and the ongoing crisis in Turkey since 2018 provide timely reminders that such concerns persist today.

However, large imbalances may also be signals for concern, or a symptom of a problem. For instance, Bernanke et al. (2011) write that international capital inflows, particularly from the GSG countries between 2003 and 2007, likely played an important role in increasing demand for U.S. government bonds and other apparently safe U.S. assets, especially mortgages, in the years leading up to the financial crisis. The GSG countries over this period are mainly China, the OPEC nations and other Asian nations. China's accumulated current account surplus was approximately 900 billion USD over the period, almost all of which was invested into U.S. assets. For OPEC it was 700 billion USD, of which 100 billion USD went directly into U.S. assets and for other Asian nations, it was 500 billion USD, of which approximately 150 billion USD went into U.S. assets.

The GSG countries purchased much of the supply of traditional safe assets, such as government bonds and high rated corporate bonds and therefore demand remained from Europe and elsewhere. This demand may have incentivised the U.S. financial services industry to engineer mortgage loans into the apparently safe, AAA-rated securities via the securitisation process. The subsequent bursting of the housing bubble and recognition that many of these securities were far riskier than had previously been recognised helped to trigger the GFC.

More generally, if market participants perceive a persistent current account deficit to be unsustainable they may look to sell the currency of the country in deficit. Such currency weakening is common, and it is a driver of normal FX volatility. However, it can also be an abrupt *adjustment* leading to sharp devaluation and excessive volatility in FX markets. In the language of the previous chapters, the FX weakening occurs through the inflation channel and the risk premium channel. Inflation expectations rise because the market perceives that the deficit country will print currency and inflate their way out of the financial liabilities it has built up, and the risk premium associated with the currency increases due to the economic turmoil. The FX price dynamics are well described by Example 5 in Sect. 3.4.5.

5.5 Consumption, Investment and Government Spending

This chapter has so far focussed on the BOP. However, the remaining components of the National Income and Product Accounts (NIPA), namely private consumption, investment and government spending may also be important in determining the FX rate. For example, in Chapter 6 I introduce the extended Mundell-Fleming model, which explicitly studies how these variables, the BOP and the real FX rate are linked together via the real interest rate in a classical macroeconomic framework. Further, in Chapter 7 I discuss how these variables may influence the nominal FX rate through their impact on inflation.

The key relationship is the NIPA identity,

$$Y = C + I + G + NX. \tag{5.4}$$

Here, Y is GDP, C is private consumption, G is government spending net of transfers, I is investment, and NX is the trade balance.

To understand where this equation comes from, consider the following. Each year the total goods and services supplied to the domestic economy is

$$\text{Supply} = \text{Total Output of Domestic Firms} + M, \tag{5.5}$$

where M denotes total imports. These goods are used as follows.

$$\text{Use} = X + I + C + G + \text{Intermediate Purchases}, \tag{5.6}$$

where X denotes total exports. I refers to goods such as plant and equipment that are produced, and also to inventories that are built up by firms. C refers to goods and services that are consumed by households. G is government expenditure. Intermediate Purchases refers to items that are purchased on the way to providing a good or service. For example, the supply of wood to the factory that makes the table would be considered an intermediate purchase.

Finally, note that GDP, Y, is defined as the total market value of *final* goods and services produced by the economy in a year. Therefore,

$$Y = \text{Total Output of Domestic Firms} - \text{Intermediate Purchases}. \tag{5.7}$$

In our example, we do not want to include both the value of the wood, and the value of the table in the GDP, because the wood was used to make the table. Hence we must subtract intermediate purchases in Eq. (5.7).

Finally, setting Eqs. (5.5) and (5.6) equal to each other, so that Supply = Use, substituting Eq. (5.7) in for Y, and setting $NX = X - M$ gives us the NIPA identity in Eq. (5.4).

5.5.1 Current Account, Savings, Investment and Twin Deficits

Let us put the NIPA identity to use. Here, we use it to show that the current account balance is a nation's savings minus it domestic investment. This result is intuitive; anything that is produced, but not consumed or invested at home is exported and therefore appears in the current account balance.

Another way to understand this relationship is via the financial account. Over the course of the chapter, we have learnt that the current and financial accounts are two sides of the same coin. Since the financial account reflects the accumulation of foreign assets, it is logical that this should be equal to the amount that a nation saves, less the amount it invests domestically. Let us derive this relationship more formally.

The Gross National Disposable Income (GNDI) of a country is defined as

$$GNDI = Y + PIB + SIB. \tag{5.8}$$

In words, it is the total income of residents, whether it is earned in the country and therefore included in GDP, Y, or earned abroad, in which case it is included in PIB or SIB. Combining this equation with the NIPA identity, Eq. (5.4), and recalling that $CAB = NX + PIB + SIB$ (see Eq. 5.1), we may write

$$GNDI = C + I + G + CAB. \tag{5.9}$$

Next, define national savings, S, as the remaining income after private consumption and government spending,

$$S = GNDI - C - G. \tag{5.10}$$

Finally, substituting Eq. (5.10) into Eq. (5.9) we have

$$CAB = S - I. \tag{5.11}$$

As expected, the current account balance reflects a nation's savings less investment.

The Current Account Balance and the FX Rate

Equation (5.11) provides further support to the view that the relationship between the current account balance and the FX rate is not clear. Note that all $CAB < 0$ tells us is that $I > S$. If $CAB < 0$ because S is low, then indeed one may suspect some profligacy that has the potential to be unsustainable and could lead to a weakening in the FX rate. However, we may also have $CAB < 0$ because I is high. If the country's investment opportunities are attractive, then both domestic and foreign residents may wish to exploit them, thereby driving up I. Investors may interpret such a situation as positive for the FX rate.

In the U.S., the persistent current account deficit has not been to the detriment of the USD, nor has it damaged its safe-haven status. I discuss this in more detail in the next feature box. However, a growing current account deficit has signalled currency devaluation in several economies such as Turkey. The key take away is that the sign of the *CAB* is insufficient to signal how sentiment towards the FX rate may evolve, and its breakdown is what matters.

The logic should be relatively clear. However, there have been misinterpretations of what a current account deficit means for a currency by prominent commentators in the past. I discuss some of the more famous examples in the next feature box.

The U.S. Current Account and the Dollar

There have been many forecasts of a collapse in the U.S. dollar that were predicated on U.S. current account deficits. Perhaps the most poignant were those just ahead of the U.S. mortgage crisis, that formed the first phase of the GFC. Below, I present two examples.

In Krugman (2007) Nobel Laureate Paul Krugman wrote:

> Almost everyone believes that the US current account deficit must eventually end, and that this end will involve dollar depreciation. However, [though] many believe that this depreciation will take place gradually...there will at some point have to be a 'Wile E. Coyote moment' a point at which expectations are revised, and the dollar drops sharply.

During the same period of time, Harvard economist and former chief economist of the IMF Kenneth Rogoff wrote:

> This is all pointing to a greatly increased risk of a fast unwinding of the U.S. current account deficit and a serious decline of the dollar. We could finally see the big kahuna hit.

As the crisis grew however, both U.S. residents and foreign investors moved money into the U.S. Prasad (2015) writes:

> From September to December 2008, U.S. securities markets had net capital inflows of half a trillion dollars, nearly all of it from private investors. This was more than three times the total net inflows into U.S. securities markets in the first eight months of that year...In contrast, many advanced economies, including [current account surplus economies] Germany and Japan, experienced overall net outflows...

Economic crises are often described as periods of *corrections*. However, with the U.S. sitting at the epicentre of the crisis, the reader may infer that the U.S. current account deficit was not something that needed to be corrected. Indeed, U.S. assets (mainly safe assets in this particular circumstance), provided the investment opportunities desired by foreign residents, and the USD duly benefited.

Finally, a separate, but related point centres on the risk premium, discussed in Chapter 1. The U.S. is able to provide the investments that foreign residents wish to hold during crisis periods (mainly U.S. safe assets). As discussed in Chapter 2, there is therefore a UIP violation in favour of the USD. This does not seem to be impacted by the U.S. current account deficit.

5.5.2 Twin Deficits

The term *twin deficits* refers to the circumstance in which the nation runs an overall current account deficit, and its government simultaneously runs a fiscal deficit. To understand how these two deficits are related to each other and to the current account, let us further decompose Eq. (5.11).

S may be broken down into government savings, S_g, and private savings, S_p, as follows.

$$S_g = T - \text{Transfers} - \text{Interest} - G, \text{ and} \tag{5.12}$$

$$S_p = GNDI + \text{Transfers} + \text{Interest} - T - C. \tag{5.13}$$

S_g can be understood as follows. T is the total tax revenue collected by the government. From this, it must pay transfer payments and interest payments on its debt to residents, and for government expenditure, G.

Similarly, S_p can be understood as follows. The income of residents is the *GNDI*, plus transfers received from the government, plus interest income received on the government debt. From this income, residents must pay their taxes, and for their consumption. The remainder is private savings, S_p.

We therefore have that

$$S_p + S_g = GNDI - C - G = S, \tag{5.14}$$

where the second equality follows from Eq. (5.10).

Finally, write $I = I_p + I_g$, where I_p denotes private investment I_p, and I_g denotes government investment.

Substituting Eqs. (5.12) and (5.13) into Eq. (5.11) we have

$$CAB = (S_p - I_p) + (S_g - I_g). \tag{5.15}$$

Here, the term $S_g - I_g$ is the fiscal deficit. The important point to note is that a fiscal deficit does not cause a current account deficit. The reason is that $S_p - I_p$ may be large enough to cover the fiscal deficit. However, all else being equal, and increasing fiscal deficit increases the current account deficit.

BOP, Currency Pegs and Triffin's Dilemma

The starting point for a country's currency to become a *reserve* currency, by which I loosely mean one that is held in large amounts by foreign central banks as part of their reserve assets, and one that is widely used in international trade, is that the issuing country must supply it to the rest of the world.

In the aftermath of the World War II, the USD proceeded to become the dominant reserve currency. However, the BOP identity, Eq. (5.2), taught us that since the current account and financial account are opposite sides of the same coin, to supply USD to the rest of the world, the U.S. was required to run current account deficits.

During this period, the Bretton Woods agreement had fixed the price of gold at 35 USD per ounce. As the U.S. proceeded to run current account deficits, USD continued to flood into the rest of the world. In his 1959 address to Congress' Joint Economic Committee, and in his influential book, Triffin (1960), economist Robert Triffin noted that the supply of USD to the rest of the world would eventually become too large relative to the U.S. gold stock at the pegged price of 35 USD per ounce. Were foreigners to convert their increasingly abundant USD holdings into gold at this price, the U.S. would be unable to meet its obligations. This would lead to a *confidence problem*, a run on the USD versus gold, and thereby a collapse of the Bretton Woods system.

Therein lay the *Triffin dilemma*; continue to run a current account deficit and collapse the Bretton Woods system, or stop running deficits, and thereby reduce global USD liquidity, which could damage international trade and the reserve status of the USD.

Triffin was prescient. The U.S continued to run current account deficits and the Bretton Woods system duly collapsed in 1971, in what became known as the *Nixon Shock*, after President Richard Nixon announced that USD would no longer be convertible to gold at the fixed rate. By the end of 1972, gold had risen to 65 USD per ounce, and then to 670 USD per ounce by September 1981.

The U.S. Secretary of the Treasury, Henry H. Fowler, summed up the feeling of the time. He wrote:

> Providing reserves and exchanges for the whole world is too much for one country and one currency to bear.

A Critique of Triffin's Dilemma

Note that the U.S. has continued to run the largest current account deficits in the world (see Fig. 5.1). However, in this time, the reserve status of the USD has progressively increased (see, for example, Prasad [2015]). Further, as discussed in Chapter 2, the U.S. was eventually described having an *exorbitant privilege* as a result. Being the world's reserve currency cannot be a privilege and a burden at the same time! This author's interpretation is therefore that Fowler's comment is not strictly correct.

The reason that Bretton Woods collapsed is because of inflation. If the rate of inflation is positive, then it is not possible to peg to a real asset indefinitely because, definitionally, its price must be increasing. I discuss inflation in detail in Chapter 7. However, at this stage, the reader may note that inflation has remained low and stable for an extended period, post Paul Volcker's chairmanship of the Federal Reserve, while the supply of USD assets to the rest of the world has notably increased. Running a current account deficit is therefore not necessarily a cause of inflation, and may not have been the cause of the collapse of Bretton Woods. However, inflation due to any reason at all may cause a currency peg to break.

5.6 The Government/Public Sector

FX investors are familiar with public sector announcements, such as an infrastructure package, a change in the taxation rate, a fiscal stimulus, an implementation of QE by the central bank, or other action impacting asset prices and the FX rate. A starting point to understand how such announcements may impact the economy is to understand the key features of the public sector accounts. The remainder of this section is to provide an introductory tour of the key fiscal statistics.

The chapters that follow then provide a framework in which to assess the impact of changes in key fiscal statistics on the FX rate. For example, Chapter 6, studies the impact of the government's contribution to aggregate demand on the FX rate when the price level is fixed in a Keynesian model of the economy. Here, I show that the *overall balance* is the corresponding fiscal statistic. Chapter 7 discusses the Fiscal Theory of the Price Level (FTPL) and how the key fiscal statistic, the primary net operating balance, studied in this chapter impacts inflation. The FX rate should then respond through the inflation channel via the mechanisms discussed in Chapter 3.

Since countries report such accounts in different ways, I focus here on the Government Finance Statistic Manual (GFSM) standard method, designed by the IMF, which provides a common language for reporting. The reader will be able to understand the format of the historical fiscal statistics provided on the website of the IMF, but will need to interpret the terms described here to find data reported in individual country's national accounts. However, in practice the reader may wish to understand the basics of government accounting to interpret how headline news may impact accounts on a forward basis, as opposed to valuing FX using historic accounting data. The emphasis here is therefore on understanding the accounting process so as to provide the depth of understanding that is necessary to make forward looking statements and I do not study any individual country's past or current fiscal accounts.

Government Net Worth

The central relationship is,

$$\Delta NW = NOB + OEF\,.$$

Here, ΔNW is the change in the net worth of the government, defined as the change in total assets less total liabilities over the year. The government's net worth is affected by the net operating balance, NOB_t, which is provided in the Statement of Government Operations (SGO). It is also affected by other economic flows, *OEF*, which is provided in the Statement of Other Economic Flows (SOEF). Over the coming subsections, I explain these variables, the SGO and the SOEF in more detail. At this stage, the reader should think of *NOB* as broadly determined by fiscal policy (taxes less expenses), and *OEF* loosely as the mark to market on existing assets. Before we define these variables, let us begin by discussing what exactly constitutes the government from an accounting standpoint.

5.6.1 What Constitutes the Government?

From a financial standpoint, the government may be thought of as consisting of two parts. The first is the general government, comprised of the central government, state governments and local governments. The second is public corporations under the control of the general government, but with separate accounting and activities. Public corporations can be financial institutions such as the central bank, and government entities (such as Fannie Mae and Freddie Mac in the U.S.), and non-financial institutions such as nationalised utility and oil companies, and others. One may look at the fiscal statistics that follow for each section of the government, or in aggregate to understand the public sector as a whole.

5.6.2 Statement of Government Operations (SGO)

The Statement of Government Operations is a flow statement with two components. The first is the *NOB*, and the second is net lending and borrowing, *NLB*. Let us study each of these in turn.

Net Operating Balance

Let R denote the revenues of the government. These come mainly from taxes, but may also contain grants and other revenues. Next, let E denote the expenses of the government, including the wages of employees, interest expenses on debt, depreciation of infrastructure such as roads and bridges, transfers as part of social benefits, among others. The net of these is a key statistic, called net operating balance *NOB*. That is,

$$NOB = R - E.$$

The reader may interpret *NOB* as government savings S_g from Eq. (5.12). To see why, note that T is similar to R, and transfers, interest, and government consumption, G, enter as expenses E.

From this equation, the reader may note that a package of tax cuts lowers the *NOB* due to lower R, and increased social transfers lower *NOB* through higher E_t. Such fiscal packages therefore directly impact the government's net worth.

Net Lending and Borrowing

The second component is transactions in financial and non-financial assets. Unlike R and E, transactions in financial and non-financial assets do not impact government net worth because they are recorded on the balance sheet at the price paid/or received. Financial assets are items such as currency and deposits, loans, equity and debt in public corporations, monetary gold and Special Drawing Rights (SDRs) held with the IMF, among others and liabilities. Non-financial assets are items such as fixed assets, like buildings and machinery, infrastructure projects, inventories, or non-produced assets such as land, or oil fields, among others. Profit and loss-making transactions will come back to impact net worth over time, but will do so via the *NOB* in the future, and the SOEF, which I discuss in the next section.

Combining the components of the SGO, we have another key statistic, the net lending and borrowing of the government, *NLB*, defined as,

$$NLB = NOB - NANA, \tag{5.16}$$

where *NANA* stands for the net acquisition of non-financial assets. The reader may think of *NANA* analogous to government investment, I_g, from Sect. 5.5.2. *NLB* is therefore similar to $S_g - I_g$, which is the government's fiscal balance.

If *NLB* is negative then the government must finance its activities. This is clearly more likely to happen if the government runs the fiscal deficit leading to a negative *NOB*. It must finance through incurring additional government debt and thereby increasing its liabilities, or through drawing down assets if it has the ability to do so. Another way of looking at *NLB* is therefore that it equals the net acquisition of financial assets minus the net incurrence of liabilities (financial and non-financial).

The reader may question why net acquiring non-financial assets appear to be treated differently to acquiring financial assets. The reason is that if the government were to acquire a corporate bond, for instance, then this would be included in *NLB* as part of its lending. However, an infrastructure project would be recorded as an asset on the balance sheet and the government would need to borrow cash funds against this asset to implement the project. It would typically do this via issuing new debt, causing a more negative *NLB*.

The final key statistic in this section is the overall balance (*OB*), defined as

$$OB = NLB - \text{Privatisation Proceeds} - \text{Policy Loans}.$$

Policy loans are loans to assist a certain sector of the economy, or to fulfil a government objective that the private sector may not satisfy. For example, student loans or social housing loans may fall into this category. In essence, the *OB* treats both privatisation proceeds and policy loans as though they were an expense to provide a more conservative measure of the *NLB*, because privatisation proceeds may be one-offs, and policy loans may not be valued based on a fair market price.

The *OB* can be thought of as the government's contribution to aggregate demand. *NLB* reflects government spending on its consumption G, plus government investment I_g. However, since privatisation proceeds are likely to be one-off, and policy loans are likely to be part of a government's economic agenda, we must further subtract these off to calculate the government's aggregate demand contribution. The next chapter discusses how the government's contribution to aggregate demand may impact the FX rate.

Finally, one may calculate the *primary* net operating balance, and the *primary* overall balance by adding back interest expenses. That is, these quantities are given by NOB + Interest Expenses and OB + Interest Expenses respectively. Since interest expenses relate to past fiscal policy, primary balances are usually used to assess a government's current fiscal policy.

Common Government Announcements

The reader should note a key difference between (i) a fiscal announcement of a change in the rate of taxation and/or social transfers, and (ii) the announcement of an infrastructure package.

(i) impacts the net worth of the government. The reason is that a change in taxation impacts revenues R, and a change in transfers impacts expenses E. Together, *NOB* is impacted. Further, the amount that the government must borrow *NLB* is impacted through Eq. (5.16).

However, (ii) does not impact net worth because infrastructure is recorded as a non-financial asset on the balance sheet. It does still impact the amount that the government must borrow *NLB* through an increase in *NANA*. Further, an infrastructure package may impact government net worth over time depending on the performance of the investment. For example, if it is able to create future increases in tax revenues R beyond its depreciation, which is recorded in E, the net worth increases over time because these items will be recorded in future values of NOB.

5.6.3 Financing the Government Budget

NLB must be financed by the sources listed in the following equation,

$$-NLB = \underbrace{\text{Central Bank} + \text{State-Owned Banks} + \text{Private Banks}}_{\text{Financial Sector}} + \text{Non-Financial Private Sector} + \underbrace{\text{Non-Residents}}_{\text{Foreign Financing}}.$$

Recall that the most comprehensive view of the government incorporates both the central bank and state-owned banks, and therefore these components would not appear in this equation if one is taking this view.

Each source of financing has potential drawbacks. If private banks and the non-financial private sector finance the general government's NLB then this may crowd out the private investments that these funds would have otherwise flowed into. Similarly, if non-residents finance the government, then this may lead to an accumulation of foreign debt, which could be the source of a currency crisis, particularly for an emerging economy. The benefits of a fiscal policy of running a negative NLB must be greater than these drawbacks.

Financing via the central bank is more interesting. Below, I discuss this approach in more detail.

Central Bank Financing of the Central Government

In the U.S., U.K. and many other countries, the finance department of the central government is referred to as the Treasury. Taking the widest definition of the government as consisting of the central government and public corporations (including, of course, the central bank), it seems strange that central bank financing of the Treasury can have any economic impact at all because the central bank is part of the broad government. Taking an analogy, if a child borrows money from her father, there can be no impact on the net worth of the family as a whole, because the transaction is made within the family. There can also be no corresponding change in the balance sheet of the family as a whole with the rest of the world because the transactions are offsetting within the family. Why then should a central bank finance the Treasury, and can such action have any economic impact at all on the private sector?

The central bank provides the Treasury with newly created cash/reserves in exchange for a government bond. The Treasury can spend this cash to pay its creditors today. It must then eventually pay back the central bank the value of the bond as it matures.[5] If it does this then the reserve balance that was created eventually returns to zero. All the central bank has therefore really done is a type of transformation—it has provided money that creditors want, cash/reserves, in place of something that they presumably did not want, a government bond.

Why would creditors not want to own the government bond? If they did, then the central bank need not be involved at all. The central government could simply sell more government bonds to meet its funding requirements. There may be liquidity, institutional, market or other reasons, in which case such an approach to financing is economically relatively benign. However, the reason that is of most concern from the point of view of the FX market is if the creditors did not want the bond because they suspect that the central government may default on it either explicitly or via inflation in the future. They may therefore prefer cash to spend today. This may lead to inflation now, and a devaluation in the FX rate through the mechanism described in Chapter 3.

Returning to the analogy, since the father-child transaction does not impact the family balance sheet with the rest of the world, the only way that the transaction may impact the economy is if it provides a signal that perhaps the daughter may spend money in a different way to her father, and that may impact prices today through the expectations channel. Similarly, although a Treasury that is financed by its central bank has no balance sheet impact when viewed as a whole, the signal that creditors and private investors do not wish to own *more* government debt may be a poignant one with inflationary market implications. In summary, the key argument is that the transaction itself has no impact, but the signal may.

Relationship with QE

Since the GFC, central banks, particularly in the developed world, have implemented QE programs that involve buying government debt in exchange for reserves created by the central bank. These purchases have not led to the excessive inflation that market participants had feared when the programs were announced. How can we reconcile this with the discussion of central bank financing of the Treasury in the previous subsection? The key difference is that QE involves the central bank purchasing government bonds from the private sector. Its aim is not to facilitate government payments. QE therefore does not necessarily provide a negative signal relating to government finances.

[5] Since central bank accounting profits are typically returned to the Treasury, it is the price paid for the government bond that the Treasury must return to the central bank, rather than the principal plus coupon payments.

In fact, in the U.S. the Federal Reserve is not permitted to participate in Treasury auctions, beyond reinvesting the proceeds of maturing securities.

Further, unlike central bank financing of the Treasury, since QE involves a public corporation trading with the private sector, it does carry a balance sheet impact for the broad government with the rest of the economy. In short, QE shortens the maturity structure of outstanding government debt. The central bank repurchases long duration bonds that were originally issued by the Treasury, in exchange for reserves. The combined balance sheet of the central government and the central bank, i.e. the broad government's balance sheet, therefore consists of fewer long duration liabilities and more reserves as liabilities.

There may be some relationship with the duration of government debt, the economy and inflation, but it appears to not be a strong one. I return to this topic when discussing the FTPL in Chapter 7.

5.6.4 Statement of Other Economic Flows (SOEF)

The SOEF accounts for changes in (i) the value and (ii) the volume of government assets and liabilities. Value and volume changes are mainly determined by market price movements and events beyond the control of the government, rather than current fiscal policy. However, market prices themselves may be influenced by other policies, such as exchange rate pegs or monetary policy. Further, the set of assets and liabilities on the balance sheet may be influenced by past policy, such as a bail out of a private sector company through the purchase of its equity.

As an example of a value change, recall that governments and central banks purchased equity in financial institutions during the GFC in so-called *bailouts*. The U.K. government purchased equity in the Royal Bank of Scotland. On March 19, 2021, the Financial Times reported that the UK government sold a proportion of this equity at a loss of 1.8 billion GBP. The U.S. government fared better at recovering its investment. On October 17, 2012, U.S. President Barack Obama tweeted "We got back every dime used to rescue the banks..."

A more common example of a value change recorded on the SOEF is the mark to market on a central bank's holding of foreign currency reserves after a change in the FX rate.

As an example of a volume change is the discovery of an oil reservoir. An example at the other end of the spectrum is a natural disaster that depletes physical assets.

5.7 Chapter Summary

- The BOP flow statement comprises of the current account (CAB), capital account (KAB) and the financial account (FAB). Of these, the current account and the financial account are usually the largest components. The current account is mostly dominated by international transactions in goods and services, and the financial account records financial transactions. The key BOP relationship is $CAB + KAB + NEO = FAB$.
- Since KAB and NEO are usually small we have $CAB \approx FAB$. That is, the current account and the financial account are two sides of the same coin. If a country runs a current account deficit by importing more goods than it exports, then it must finance this by net selling assets to the current account surplus country.
- The surplus country buys these assets as claims to the debtor country's future production. The experience of the U.S. and the U.K. proves that running current account deficits is not necessarily negative for the currency. However, if demand for deficit country assets falls, perhaps because it becomes apparent that there will be inflation and therefore the assets will provide fewer goods in return, there may be a negative impact on the currency.
- The IMF presents the BOP in standard form and in analytical form. FX investors searching for unsustainable current accounts and the potential for an FX devaluation should use the analytical form. This separates reserve assets, IMF loans and exceptional financing from the FAB. It therefore provides investors with a clearer idea of whether a negative CAB is willingly financed by the private sector.
- The NIIP is a stock statement. It logs the accrual of the FAB over time, plus valuation and volume changes. In short, the NIIP provides the net borrowing or lending of a country with the rest of the world.
- The NIPA identity decomposes Y into C, I, G, and NX. This identity can be further used to show that $CAB = S - I$. It can also be used to derive the so-called *twin deficits* relationship that says that $CAB = (S_p - I_p) + (S_g - I_g)$.
- The public sector consists of the central government and public corporations, such as the central bank. The GFSM provided by the IMF records the key fiscal statistics that may impact the economy and the FX rate.
- The net worth of the government is impacted by its net operating balance (fiscal policy) and other economic flows (mark to market on investments). Net lending and borrowing measures the government's fiscal surplus or

deficit. The overall balance is a measure of the government's contribution to aggregate demand.

- The central bank may finance the Treasury. However, this can be a sign of inflation to come and thereby a weakening in the FX rate. Although the financing process appears similar to QE, there are important subtle differences. Central bank financing of the Treasury does not change the maturity structure of government debt, however QE does.

References

Bernanke, B. S. (2005). *The global saving glut and the US current account deficit*. Published on website of the FRB.

Bernanke, B. S., Bertaut, C., DeMarco, L. P., & Kamin, S. (2011). International Capital Flows and the Returns to Safe Assets in the United States, 2003–2007. *FRB International Finance Discussion, 15*, 13–26.

Krugman, P. (2007). Will there be a dollar crisis? *Economic Policy, 22*(51), 435–467.

Prasad, E. (2015). *The dollar trap: How the U.S. dollar tightened its grip on global finance*. Princeton University Press.

Shambaugh, J. C. (2012). The Euro's three crises. *Brooking's Papers on Economic Activity, 43*, 157–231.

Triffin, R. (1960). *Gold and the dollar crisis*. Yale University Press.

6

The Mundell-Fleming Model and the Impossible Trinity

This chapter builds an understanding of how changes in the important macroeconomic aggregates studied over the previous chapter may impact the FX rate. The seminal Mundell-Fleming model provides a starting point.

The original model was formulated by economists Robert Mundell[1] and Marcus Fleming in the early 1960s. However, it takes the real interest rate as set exogenously, because it focuses on a small open economy where the interest rate is influenced by the rest of the world. FX investors focus a great deal of attention on the differentials in the real interest rates across countries, and therefore this chapter follows the approach of Mankiw (2007) and presents an extension of the model, adapted for a large open economy with control over its own monetary policy, whereby the central bank can influence the real interest rate. The model ties together the macroeconomic aggregates that we discussed in the previous chapter, namely private consumption, government spending and investment, and links them to the real exchange rate via the BOP relationship, Eq. (5.2).

The seminal IS-LM (Investment Savings-Liquidity Money) model of macroeconomic balance, developed by Hicks (1937), interpreted the work in Keynes (1936) in Keynes' Magnus opus. It has been a workhorse of practical macroeconomic analysis and policy making over an extended period of time. The Mundell-Fleming model and its extension presented in this chapter is an open economy extension of the IS-LM model. It can therefore be thought

[1] Robert Mundell was awarded the 1999 Nobel Prize for his work on open-economy macroeconomics.

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_6

of as a Keynesian representation of the macroeconomy, with the additional contribution being that it incorporates the international macroeconomic aggregate, the BOP. The IS-LM approach has well-known shortcomings, with perhaps the most relevant for investors that it is static and does not incorporate expectations, thereby falling short of the Lucas critique (Lucas, 1976). These shortcomings then pass into the Mundell-Fleming model. I discuss these in more detail over the chapter and argue that the model nevertheless provides the reader with a useful starting point to consider how the commonly cited macroeconomic aggregates interact with each other and the FX rate.

This chapter also studies the *impossible trinity* or Mundell-Fleming *trilemma*, which states that an economy can choose two conditions from (i) a fixed exchange rate, (ii) free movement of capital and (iii) independent monetary policy, but cannot choose all three. I show how the impossible trinity idea can be proven via the extended Mundell-Fleming model.

Finally, the reader may note that the approach presented in this chapter assumes that the price level remains fixed and therefore does not address the inflationary implications of changes in macroeconomic variables, and the FX response via the inflation channel. Chapter 7 extends this model and introduces others to include inflation.

6.1 Basic Equations of the Model

There are three core equations that form the model. The first focusses on the market for goods, the second on the money market and the third on the balance of payments. In short, the model solves for the interest rate such that the goods market clears, and the exchange rate such that the BOP identity, Eq. (5.2), holds.

I do not specify the precise functional forms or estimate parameters in the model to provide quantitative estimates of the exchange rate. In practice, the Mundell-Fleming approach is not often used to provide such estimates. Instead, the model is used to understand the directional (static) impact of changes in fiscal, trade and monetary aggregates on the exchange rate. I explore these ideas later in the chapter. Let us begin.

6.1.1 The Goods Market

The first equation is the NIPA identity, Eq. (5.4), but with the addition of some dependencies,

$$Y = C(Y - T) + I(r_d^{\mathrm{real}}) + G + NX(e). \tag{6.1}$$

The meanings of the parameters and variables in this equation, and the remaining functions and variables used in this chapter are recapitulated in Tables 6.1 and 6.2, respectively.

Recall from the discussion in Sect. 5.5 that Eq. (6.1) is an accounting identity. Therefore, the only modelling that we have done at this stage is in the assumptions for the dependencies of the variables on other quantities. For example, the model only analyses the dependence of I on r_d^{real}. In reality, I may depend on the level of confidence in the economy, the political and legal environment, the term structure of r_d^{real}, or numerous other factors. The remaining assumptions made in Eq. (6.1) will become clear over the remainder of the chapter.

Table 6.1 Model parameters

r_d^{real}	Domestic real rate of interest
T	Level of taxation
e	Real exchange rate. Number of foreign goods that can be purchased per unit of domestic good
	A high value of e implies a strong domestic currency

Table 6.2 Model variables and assumed functional forms

Y	Output or GDP
T	Taxes net of transfers
$C(Y - T)$	Private consumption, consisting of goods and services bought by households, as a function of after tax income, $Y - T$
$I(r_d^{\mathrm{real}})$	Investment, expressed a function of r_d^{real}
G	Government consumption
$NX(e)$	Net exports, $X - M$. Positive if the value of exported goods exceeds the value of imported goods, as a function of the real FX rate, e
M	Level of money supply. Fixed by the central bank
P	Price level
$L(r_d^{\mathrm{real}}, Y)$	Level of demand for money, as a function of Y and r_d^{real}
$CF(r_d^{\mathrm{real}})$	Capital flow. Positive if the value of capital lent abroad exceeds the value of capital borrowed from abroad, as a function of r_d^{real}

We will not study the precise functional forms of $I(r_d^{\text{real}})$ or $NX(e)$. We require only the intuitive relations that $I(r_d^{\text{real}})$ depends negatively on r_d^{real}, $\partial I(r_d^{\text{real}})/\partial r_d^{\text{real}} < 0$, and that $NX(e)$ depends negatively on e, $\partial NX(e)/\partial e < 0$. In words, we take as given that a higher domestic real interest rate (and hence higher borrowing costs) lead to lower levels of investment by firms, and that a stronger domestic currency (and hence a higher real FX rate) leads to lower exports.

Next, the model assumes that the population's marginal propensity to consume is less than 1. That is, for every 1 USD of income, the population consume less than 1 USD, so that $\partial C(Y - T)/\partial Y < 1$.

6.1.2 The Money Market

With Eq. (6.1) in place, we are now ready to move on to our second equation, which relates to the money markets,

$$\frac{M}{P} = L(r_d^{\text{real}}, Y). \tag{6.2}$$

Here, M is the quantity of money supplied. It should be thought of as the quantity of the stock of assets that are used for transactions. This quantity is difficult to define. For example, the Federal Reserve publishes measures of money each month.[2] However, which measure should be included is a debate that proceeds beyond the scope of this chapter. For our purposes, the reader should consider M to mean money held by consumers to be used in transactions, such as in currency, demand deposits and other easily accessible savings. P is the price level. M/P is therefore the *real* money supply.

$L(r_d^{\text{real}}, Y)$ denotes the money demand. The model assumes that the real money demand depends only on r_d^{real} and Y.[3] Again, we do not require the explicit functional dependence of $L(r_d^{\text{real}}, Y)$ on r_d^{real} and Y. We only use the following two intuitive relationships.

The first is that $L(r_d^{\text{real}}, Y)$ must depend positively on Y, $\partial L(r_d^{\text{real}}, Y)/\partial Y > 0$. The reason is that, when income is high, expenditure is also high. People engage in more transactions that require the use of money. Thus, greater income creates greater money demand.

[2] Federal Reserve, Money Stock Measures is available and updated monthly at https://www.federalreserve.gov/releases/h6/current/default.htm.

[3] The Mundell-Fleming model assumes sticky prices. I discuss this and inflation in Chapter 7.

The second is that $L(r_d^{\text{real}}, Y)$ must depend negatively on r_d^{real} so that $\partial L(r_d^{\text{real}}, Y)/\partial r_d^{\text{real}} < 0$. The reason is that high-interest rates increase the cost of holding money, rather than investing it in interest bearing account or other asset.[4]

Equation (6.2) therefore says that the real money supply, M/P, must equal the real money demand $L(r_d^{\text{real}}, Y)$.

6.1.3 The Balance of Payments

The third and final equation that completes the model equates capital flows and trade balances,

$$NX(e) = CF(r_d^{\text{real}}). \tag{6.3}$$

Here, $NX(e)$ is the trade balance, $X - M$. Applying Eqs. (5.1) and (5.2) from our analysis of the BOP, we see that

$$NX(e) \equiv X - M = FAB - (KAB + SIB + PIB). \tag{6.4}$$

Therefore, by Eq. (6.3),

$$CF(r_d^{\text{real}}) = FAB - (KAB + SIB + PIB). \tag{6.5}$$

That is, $CF(r_d^{\text{real}})$ is simply the net capital flow.

The intuition underlying Eq. (6.3) can be understood as follows. First, consider again the example of a Japanese car exported to the U.K. from Sect. 5.1.1. In the first scenario, the trade balance from the perspective of the U.K. was $NX(e) = -15{,}000$ GBP, and this was invested into U.K. assets. Since $KAB = SIB = PIB = 0$, we have that $CF(r_d^{\text{real}}) = FAB = -15{,}000$ GBP. That is, the U.K. incurred liabilities of 15,000 GBP. Equation (6.3) holds, as expected.

In scenario 2, the trade balance is zero, $NX(e) = 0$ GBP, because exports equal imports. The financial account balance is also zero, because no investments are made, $FAB = 0$ GBP. Therefore, $CF(r_d^{\text{real}}) = FAB = 0$ GBP, since $KAB = SIB = PIB = 0$ and so, again, Eq. (6.3) holds.

[4] The money demand function may seem confusing in a world in which financial innovation progressively decreases the need for traditional money. With little need for money, why should households demand it at all? Modern macroeconomic theory, such as the Fiscal Theory of the Price Level, removes the need for the money demand function to analyse inflation. I discuss this topic in Chapter 7.

Next, consider an unrelated transaction in which a resident of Japan works in the U.K. and is paid 20,000 GBP for her work. She invests this into shares of a U.K. public company. In this case, the trade balance is zero, $NX(e) = 0$ GBP, because no goods were exported. However, from the perspective of the U.K., $FAB = -30{,}000$ GBP due to the incurrence of liabilities to a foreign resident, and $PIB = -30{,}000$ GBP due to the wages she was paid. Applying Eq. (6.5) we have that $CF(r_d^{\text{real}}) = FAB - PIB = 0$ GBP, since $KAB = SIB = 0$. Again, Eq. (6.3) holds as expected.

The final important points to note are the dependences of $CF(r_d^{\text{real}})$ on r_d^{real}, and of $NX(e)$ on e. $CF(r_d^{\text{real}})$ is the amount of capital lent abroad minus the amount of capital lent to the domestic economy from abroad. If r_d^{real} is high then investors are incentivised to invest in the domestic economy and less capital flows abroad. Hence, $CF(r_d^{\text{real}})$ depends negatively on r_d^{real}, $\partial CF(r_d^{\text{real}})/\partial r_d^{\text{real}} < 0$.[5]

Since $NX(e)$ is the trade balance, the value of exports minus the value of imports, we intuit that this dependence is negative, $\partial NX(e)/\partial e < 0$. That is, the trade balance decreases if the real value of the domestic currency, in terms of the foreign currency, increases because that country's goods become more expensive for foreigners to purchase.

The assumption that net exports depend on the real exchange rate and capital flows depend on the interest is what will ultimately link interest rates and exchange rates.

6.2 The Extended Mundell-Fleming Model

With the three model equations in place, I now introduce the model. There are three steps. First, I calculate the level of the interest rate that is consistent with Eq. (6.2) and the level of output. This is the LM curve. Second, I repeat this for Eq. (6.1). This is the IS curve. Finally, I use this interest rate and Eq. (6.3) to extract the level of the FX rate.

[5] Perhaps better expressed, CF should depend on the interest rate differential between home and abroad $r_d^{\text{real}} - r_f^{\text{real}}$. However, in the analysis of this chapter, we do not allow r_f^{real} to change, and so suppress it in the notation here.

6.2.1 The LM Curve

The first step is to draw the LM curve. This curve plots the level of r_d^{real} that must prevail in the economy for each level of Y for the money markets to clear. That is, the LM curve is drawn to satisfy Eq. (6.2).

The LM curve is shown in Fig. 6.1. It is intuitive that it is upward sloping. The reason is that if economic output Y increases, then the demand for money increases. Since the real supply of money is fixed, r_d^{real} must rise to lower the demand back down to M/P.

Readers familiar with calculating total derivatives may inspect the following derivation of the LM curve. A refresher on this topic is provided in Appendix C.

Deriving the LM Curve (Optional)
The change in $L(r_d^{\mathrm{real}}, Y)$ for a change in r_d^{real} and Y is given to first order by

$$\mathrm{d}L = \frac{\partial L}{\partial r_d^{\mathrm{real}}}\mathrm{d}r_d^{\mathrm{real}} + \frac{\partial L}{\partial Y}\mathrm{d}Y. \tag{6.6}$$

If the real money supply, M/P, is fixed, then $\mathrm{d}L = 0$. Equation (6.6) can then be written as,

$$\frac{\mathrm{d}r_d^{\mathrm{real}}}{\mathrm{d}Y} = -\frac{\frac{\partial L}{\partial Y}}{\frac{\partial L}{\partial r_d^{\mathrm{real}}}} > 0. \tag{6.7}$$

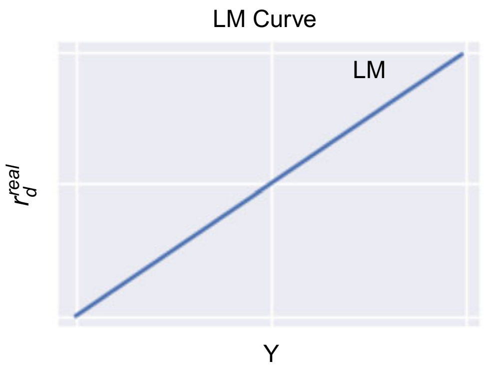


Fig. 6.1 The LM curve is upward sloping because an increase in Y leads to an increase in the money demand function, which is then offset by a decrease in r_d^{real}

As discussed in the previous section, $\frac{\partial L}{\partial r_d^{\mathrm{real}}}$ is negative and $\frac{\partial L}{\partial Y}$ is positive, making $\mathrm{d}r_d^{\mathrm{real}}/\mathrm{d}Y$ positive, to provide an upward sloping LM curve.

6.2.2 The IS Curve

Next, I follow a similar process to draw the IS curve. This curve plots the level of r_d^{real} that must prevail in the economy for each level of Y in order for the goods market to clear. That is, the IS curve is drawn to satisfy Eqs. (6.1) and (6.3).

Substituting (6.3) into (6.1) gives

$$Y = C(Y - T) + I(r_d^{\mathrm{real}}) + G + CF(r_d^{\mathrm{real}}). \tag{6.8}$$

Figure 6.2 shows that the IS curve is downward sloping. To understand why, suppose that Y rises by 1 unit. The right-hand side of Eq. (6.8) must therefore rise by 1 unit. However, we assumed that marginal propensity to consume is less than 1, and so $C(Y - T)$ rises by less than 1 unit. Since G does not depend on r_d^{real}, the shortfall must be filled by $I(r_d^{\mathrm{real}}) + CF(r_d^{\mathrm{real}})$ rising. This can only happen if r_d^{real} falls, because such a fall stimulates investment and increases capital outflows for the reasons discussed in Sect. 6.1. That is, rises in Y are associated with falls in r_d^{real}, resulting in a downward sloping IS curve. Again, a more formal derivation of the IS curve is provided below.

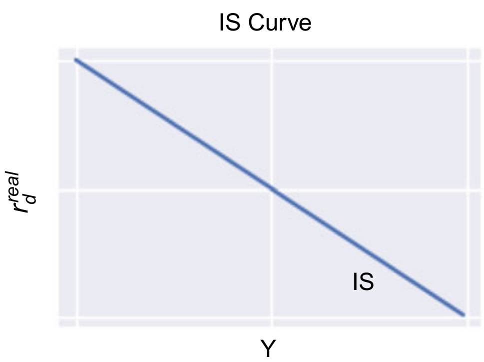


Fig. 6.2 The IS curve is downward sloping. An increase in Y cannot be accounted for with an increase in $C(Y-T)$, because households' marginal propensity to consume is less than 1. Therefore, r_d^{real} must fall to boost $I(r_d^{\mathrm{real}})$ and $CF(r_d^{\mathrm{real}})$

Deriving the IS Curve (Optional)
First, rewrite Eq. (6.8) as

$$f(Y, r_d^{\text{real}}) = 0,$$

where

$$f(Y, r_d^{\text{real}}) \equiv Y - C(Y - T) - I(r_d^{\text{real}}) - G - CF(r_d^{\text{real}}). \tag{6.9}$$

The change in $f(Y, r_d^{\text{real}})$ for a change in r_d^{real} and Y is given by

$$\mathrm{d}f = \frac{\partial f}{\partial r_d^{\text{real}}}\mathrm{d}r_d^{\text{real}} + \frac{\partial f}{\partial Y}\mathrm{d}Y, \tag{6.10}$$

to first order. Taking partial derivatives of Eq. (6.9) we find that,

$$\frac{\partial f}{\partial r_d^{\text{real}}} = -\frac{\partial I}{\partial r_d^{\text{real}}} - \frac{\partial CF}{\partial r_d^{\text{real}}}, \text{ and} \tag{6.11}$$

$$\frac{\partial f}{\partial Y} = 1 - \frac{\partial C}{\partial Y}. \tag{6.12}$$

Finally, substituting Eqs. (6.11) and (6.12) into Eq. (6.10), and noting that $\mathrm{d}f = 0$, we have,

$$\frac{\mathrm{d}r_d^{\text{real}}}{\mathrm{d}Y} = \frac{1 - \frac{\partial C}{\partial Y}}{\frac{\partial I}{\partial r_d^{\text{real}}} + \frac{\partial CF}{\partial r_d^{\text{real}}}} < 0.$$

Since both $I(r_d^{\text{real}})$ and $CF(r_d^{\text{real}})$ depend negatively on r_d^{real}, the denominator is negative, and since $\partial C/\partial Y < 1$, the numerator is positive. The slope of the IS curve, $\mathrm{d}r_d^{\text{real}}/\mathrm{d}Y$, is therefore negative.

6.2.3 Equilibrium and Analysis

Figure 6.3 provides a visualisation of how the model determines the exchange rate. Let us discuss each of the charts in turn.

The upper-left panel shows the IS and LM curves. To recapitulate, the LM curve shows all the combinations of real interest rate, and total output at which the money markets clear in that money supply equals money demand.

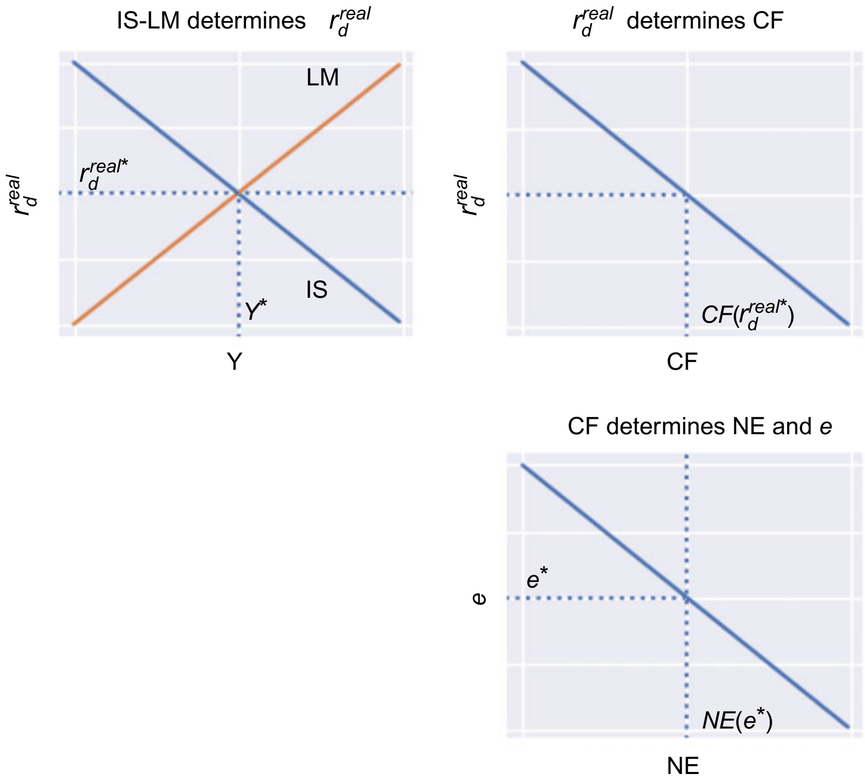


Fig. 6.3 The upper-left panel shows that the intersection of the IS and LM curves determines r_d^{real} such that the goods and money markets are clear. Call this $r_d^{\text{real}*}$. Since $CF(r_d^{\text{real}})$ depends only on r_d^{real}, the capital outflow is $CF(r_d^{\text{real}*})$. This is shown in the upper-right figure. Finally, the BOP identity tells us that $NE(e) = CF(r_d^{\text{real}})$. Therefore, the real exchange rate consistent with this equilibrium is e such that $NE(e) = CF(r_d^{\text{real}*})$. We can call this e^*

The IS curve shows all the combinations of real interest rate and total output at which the goods market clears. The intersection of these curves provides us with the economic equilibrium in which both the money markets and goods markets have cleared. Let us call this interest rate $r_d^{\text{real}*}$.

Next, the upper-right panel shows $CF(r_d^{\text{real}})$ plotted as a function of r_d^{real}. Recall that it is downward sloping, because at high levels of r_d^{real}, the economy attracts capital, and vice versa at low levels of r_d^{real}. Since the IS-LM intersection has determined $r_d^{\text{real}*}$, we know that the net capital outflow is $CF(r_d^{\text{real}*})$.

Finally, since the BOP identity, $NE(e) = CF(r_d^{\text{real}})$ must hold, we know that the exchange rate must be e^*, where e^* is determined such that $NE(e^*) = CF(r_d^{\text{real}*})$. This is shown in the bottom right panel.

6.2.4 Monetary Policy and the LM Curve

If the central bank announces a reduction in the interest rate, the currency typically weakens. Let us understand this in the context of the model.

The model assumes that the price level P is fixed. I return to changing price levels in the IS-LM model and other models of inflation in Chapter 7. With fixed P, an increase in the money supply, M, increases M/P. To maintain

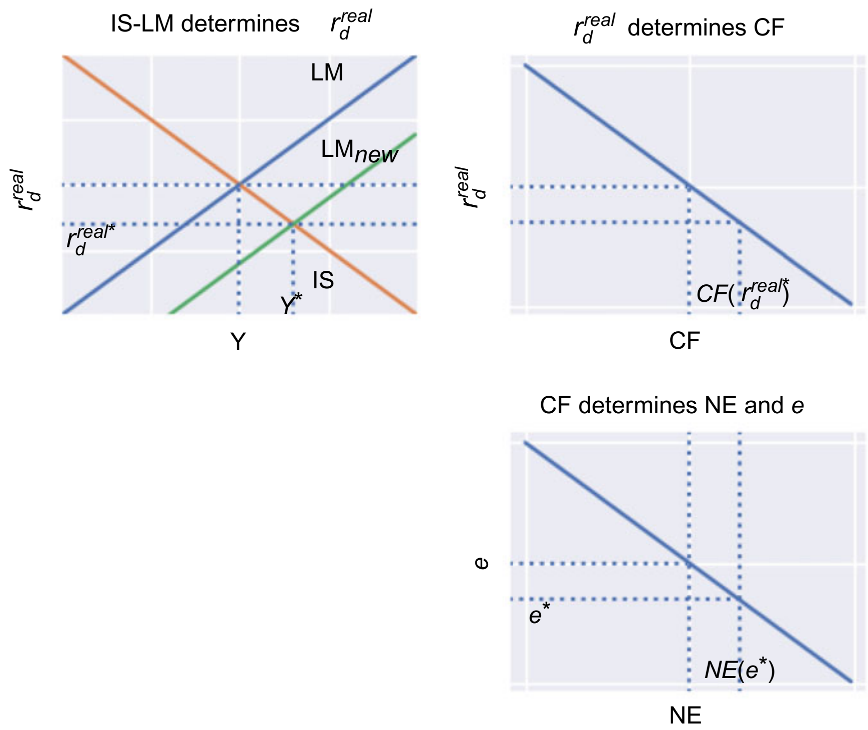


Fig. 6.4 The upper-left panel shows that the LM curve has shifted to the right in response to an increase in M. The intersection of the IS and LM curves is now at a lower level of $r_d^{\text{real}*}$. Tracing across to the upper-right panel we see that the lower $r_d^{\text{real}*}$ leads to a higher level of $CF(r_d^{\text{real}*})$, as investors prefer to withdraw portfolio investments from the domestic economy in response to the lower interest rate. Finally, the BOP identity tells us that $NE(e^*) = CF(r_d^{\text{real}*})$. The e^* is lower than previously, due to investors selling the domestic currency as they move investments abroad. Finally, note that Y^*, the level of economic output, is higher than previously

equilibrium in the money market, the LM curve must therefore shift to the right. The reason is that, for every given level of r_d^{real}, Y must be higher for the money demand to match the increase in money supply (recall Eq. [6.2]).

Figure 6.4 shows the resulting equilibrium. The IS and LM curves now intersect at a lower level of $r_d^{\text{real}*}$. This increases the net capital flowing out of the domestic economy $CF(r_d^{\text{real}})$, as investors seek higher real returns abroad. As international investors sell the domestic currency, it weakens, and this results in cheaper domestic goods and thereby a higher trade balance $NX(e^*)$.

In the U.S., such an increase in the money supply could be enacted by the Federal Reserve's conducting of open market operations, in which the Federal Reserve purchase Treasury bills from the open market, thereby increasing the supply of money, and pushing the interest rate lower. The model allows us to understand how monetary policy can impact the FX rate.

A final noteworthy point is that Y^*, the equilibrium level of economic output, is higher than previously. That is the economy has been stimulated, which may result in higher employment. Indeed, this may have been the reason that the central bank may increase the money supply, and it is the reason that such action is often referred to as *monetary stimulus*.

However, the crucial assumption is that P is fixed. If such actions lead to inflation, then this stimulative effect may be only temporary, because P may rise to leave the real money supply, M/P unchanged. I return to this topic in detail in Chapter 7.

Relationship with the Fixed Real FX Rate Model

The reader may question why an FX investor should use the more complicated extended Mundell-Fleming model discussed here, over the relatively straightforward fixed real FX rate model of Chapter 3. After all, both models provided the same implication that a fall (rise) in the real interest rate instigated by the central bank leads to a fall (rise) in the real FX rate (recall the example in Sect. 3.4.3). The reason is that each model is helpful in a different context.

Unlike the model of Chapter 3, the extended Mundell-Fleming model is not silent on the impact of a change in the real interest rate on net exports and on economic output. A macro FX investor may note that policy makers may wish to boost growth and may wish to do so because of a lagging export sector. The investor may therefore use the extended Mundell-Fleming approach to understand how such a scenario may exhibit itself.

Similarly, the extended Mundell-Fleming model is silent on the long-term real FX rate and does not impose that it is fixed, stable or follows PPP as we did in Chapter 3. It also does not directly discuss the CIP relationship. Therefore, the approach in Chapter 3 is more suitable for FX relative value investors and arbitrageurs looking to exploit CIP and differing prices in real goods across countries.

Further, the reader may note that the two approaches are not inconsistent with one another. We have already seen that the extended Mundell-Fleming model does not rule out a stable long-run real FX rate, and we shall see later in the chapter that the extended Mundell-Fleming model implies the CIP relationship.

6.2.5 Fiscal Policy and the IS Curve

Next, consider the impact of a fiscal expansion. If the government increases purchases, G rises, or cuts taxes, T, so that $C(Y - T)$ rises, then the IS curve must shift to the right. The reason is that Y must be higher at every level of r_d^{real} to maintain equilibrium in the goods market. There is no change in the LM curve.

Figure 6.5 shows the resulting equilibrium. The IS and LM curves now intersect at a higher level of $r_d^{\text{real}*}$. This decreases the net capital flowing out of the domestic economy $CF(r_d^{\text{real}})$, as investors seek higher real returns at home. As investors buy the domestic currency, it strengthens, and this results in more expensive goods and thereby a lower trade balance $NX(e^*)$.

A noteworthy point is that Y^*, the equilibrium level of economic output, is higher than previously. That is the economy has been stimulated which may result in higher employment. Indeed, this may have been the reason that the government may increase spending, or cut taxes, and it is the approach that was advocated by Keynes. This is the reason that such action is often referred to as a *Keynesian fiscal stimulus*.

6.2.6 Unaffordable Fiscal Stimulus

The reader may be sceptical that a fiscal expansion always strengthens the FX rate, as predicted by the extended Mundell-Fleming model. In particular, how may the FX market react if it perceives that the fiscal expansion will decrease the net worth of the government over time because the stimulus is unproductive in that it will not increase the government's future NOB_t? How then will the government repay the increase in debt issuance (more negative

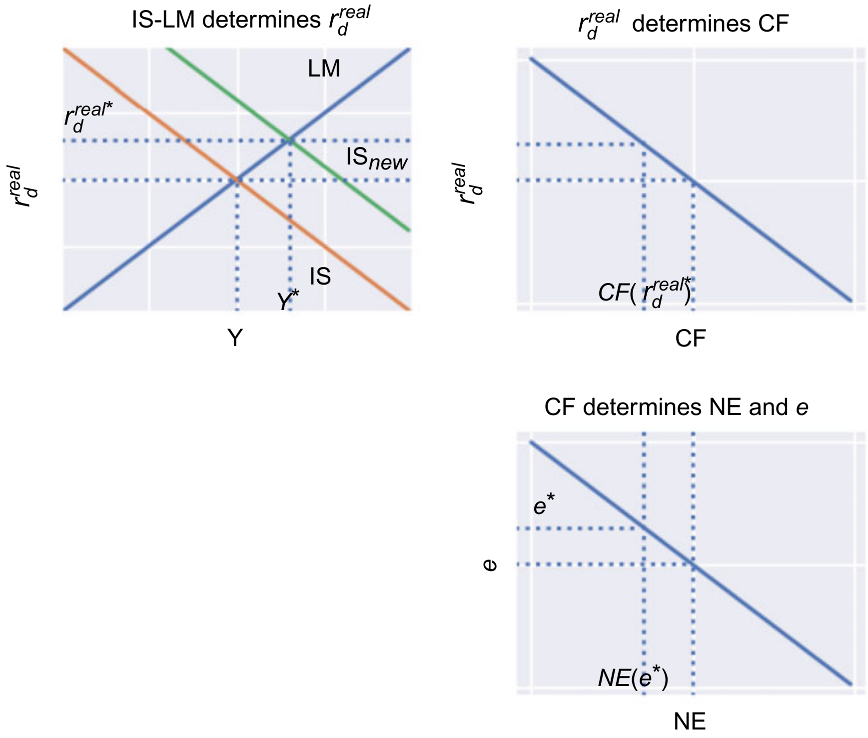


Fig. 6.5 The upper-left panel shows that the IS curve has shifted to the right in response to an increase in *G*, or a decrease in *T* leading to an increase in $C(Y - T)$. The intersection of the IS and LM curves is now at a higher level of $r_d^{\text{real}*}$. Tracing across to the upper-right panel we see that the higher $r_d^{\text{real}*}$ leads to a lower level of $CF(r_d^{\text{real}*})$, as investors prefer to increase portfolio investments into the domestic economy in response to the higher interest rate. Finally, the BOP identity tells us that $NE(e^*) = CF(r_d^{\text{real}*})$. The e^* is higher than previously, due to investors buying the domestic currency. Finally, note that Y^*, the level of economic output, is higher than previously

NLB_t) in the future? One of two things must happen—either the government must raise taxes in the future, or it must continue to roll over its debt.

Ricardian Fiscal Authority

Consumers may expect that the government will raise taxes in the future. They may therefore cut their consumption C today and save by purchasing the additional government bonds that are issued. Indeed, since the government must raise future taxes by the exact face value of the additional government bonds at maturity, by purchasing the additional debt today consumers will be left with exactly what they need to pay the future taxes.

The model presented here assumed that C depends on current T, rather than future T. Not including expectations is a potential serious shortcoming of the model. If we append this effect to the model, the cut in C today undoes the shift to the right in the IS curve in Sect. 6.2.5, the FX rate does not rise and there is no stimulative impact on the economy.

This is called *Ricardian Equivalence*, after nineteenth-century economist David Ricardo. Further, if the government spending program is inefficient in some way, it may have a negative impact on the economy and the FX rate.

Non Ricardian Fiscal Authority

Alternatively, investors may expect a so-called non-Ricardian approach in that the government will avoid tax rises by attempting to roll over its additional debt. In this case C need not change today.

If there is no known plan to raise revenue to repay the additional debt, then investors may price in the risk of default through higher inflation in the future. They may expect a monetary expansion as the central bank finances the central government, and therefore a shift to the right in the LM curve, just as in Sect. 6.2.4. This lowers the real interest rate and the FX rate.

6.3 The Impossible Trinity

The *Impossible Trinity* or *Mundell-Fleming Trilemma* is a restriction on the economic policy that a government may pursue. It argues that the government may not have all three of (i) monetary autonomy, (ii) free capital flows and (iii) a fixed FX rate. It may choose two of these, but in doing so it sacrifices the third. Figure 6.6 provides a visualisation of this idea.

Proof via the Law of One Price

The proof of the impossible trinity is straightforward in the special case where the FX rate is fixed between two nations where there exist nominal default free bond investments, because it follows from the law of one price.

To see this, suppose that the foreign and domestic economies fix the nominal FX rate and that free capital flows are allowed. That is, we are on side B of the triangle. It is then clear that the two nations cannot have monetary autonomy. If, for example, that the domestic central bank lowered interest rates below the level of their foreign counterpart, then the following arbitrage is possible.

Suppose that the DOM nominal interest rate is 0%, the FOR nominal interest rate is 3% and that the FOR-DOM FX rate were fixed at 1.0. An arbitrageur would first borrow 1 DOM unit of currency, and sell it in the FX market to buy 1 FOR unit. She may invest this FOR in foreign government

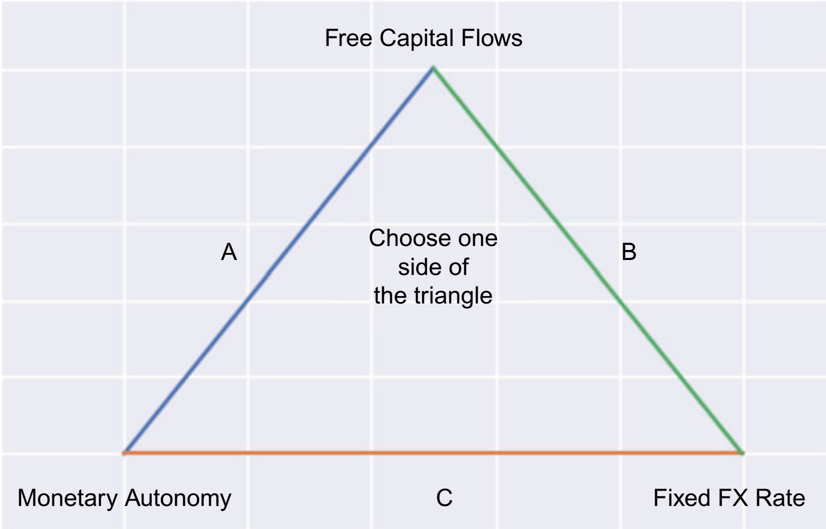


Fig. 6.6 A government may choose one side of the triangle. If, for example, it chooses policy A, so that it has monetary authority and free capital flows, then it cannot simultaneously implement a fixed FX rate. If it chooses policy B, so that it has free capital flows and a fixed FX rate, then it cannot also have monetary autonomy. Finally, if it chooses C, so that it has monetary autonomy and a fixed FX rate, then it cannot simultaneously have free capital flows

bonds to obtain 1.03 DOM in 1 year. She is able to do this because capital flows are free. Finally, she sells 1.03 FOR to obtain 1.03 DOM. She uses 1.0 DOM units to pay back her debt, thereby locking in a risk-free profit of 0.03 DOM. Since the profit is risk free, she and other arbitrageurs continue to exploit this strategy as long as it is possible to do so. The reader may have already noted that this is simply the CIP relationship from Chapter 2.

Each time an arbitrageur implements this strategy, it puts buying pressure on the FOR-DOM FX rate, until the point where there are no private buyers left, and the domestic central bank's reserves of the FOR currency are exhausted. At this point the FX peg breaks.

This example shows that side B is only possible if the two nations were to relinquish monetary autonomy and set their interest rates equal to each other to remove the arbitrage. It also shows that the nations may pursue side A of the triangle, only if the FX rate is not fixed.

Finally, to understand side C, note that the arbitrage is not necessarily possible if there are restrictions on capital mobility, because it may not then be possible for the arbitrageur to borrow in the foreign country to invest in the domestic bond. Therefore, it may not be impossible to maintain autonomous monetary policies with a fixed FX rate.

Proof via the Extended Mundell-Fleming Model and the BOP

Here, I prove the impossible trinity in the context of the extended Mundell-Fleming model developed in this chapter. The advantage of this approach is that it does not rely on the law of one price, and thereby does not require the two nation's bonds to be perfect substitutes for each other.

Let us again begin at side B of the triangle of Fig. 6.6; the FX rate is fixed and there are free capital flows. Suppose that the monetary authority increases the money supply with the objective of lowering the interest rate. Figure 6.7 shows the behaviour of the IS and LM curves, $CF(r_d^{\text{real}})$ and $NX(e)$?

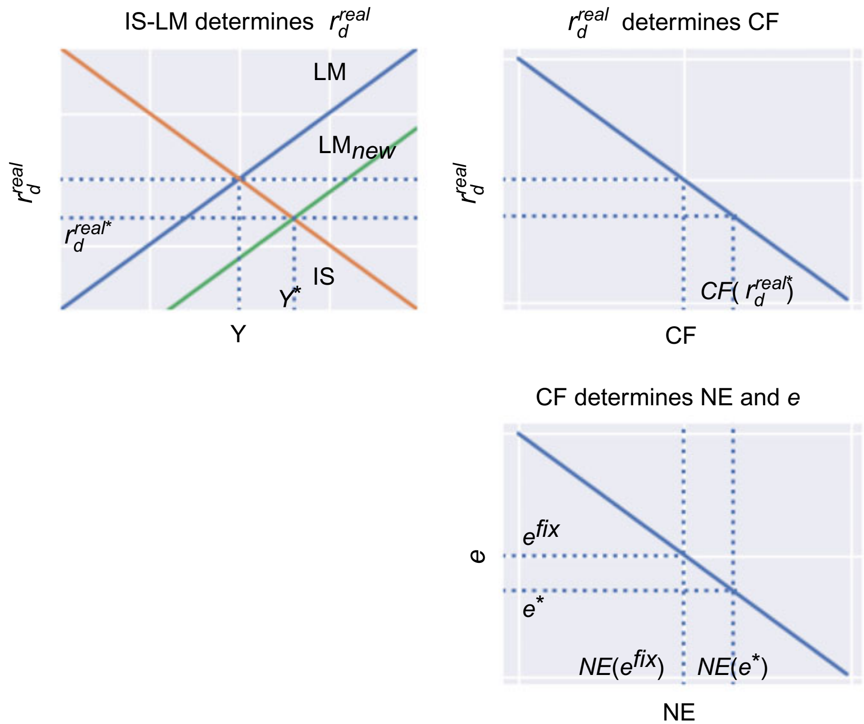


Fig. 6.7 The upper right panel shows that the LM curve shifts to the right during a monetary expansion. This lowers the interest rate to $r_d^{\text{real}*}$. The lower interest rate leads to an increase in capital outflows to $CF(r_d^{\text{real}*})$. However, if the exchange rate is fixed at e^{fix}, and cannot cheapen, net exports are stuck at $NE(e^{\text{fix}})$, which is less than e^*. The demand for capital outflow/foreign investments puts pressure on the domestic currency to devalue

The upper-left panel shows that, as in Sect. 6.2.3, an increase in the money supply shifts the LM curve to the right. This lowers the interest rate from r_d^{real} to $r_d^{\text{real}*}$. The upper right panel shows that the lower domestic interest rate leads to a desire for capital outflows, as investors wish to invest in relatively higher interest rate assets abroad. The upper right panel shows that the desired level rises from $CF(r_d^{\text{real}})$ to $CF(r_d^{\text{real}*})$. However, the FX rate is fixed and so it cannot depreciate to allow net exports to rise to match the desired increase in capital outflows and therefore, initially at least, remains at $NE(e^{\text{fix}})$.

We now have an unsustainable situation because $CF(r_d^{\text{real}*}) > NE(e^{\text{fix}})$. Since $NE = CF$ is an identity (see Eq. [6.3]), something has to give to make NE move higher, or CF lower such that the equality is met. Let us consider the financial flows that will take place.

The exporters that are receiving foreign currency for their exports will quickly find investors who wish to hold foreign currency deposits and foreign financial assets buying the foreign currency from them in exchange for the domestic currency. Since $CF(r_d^{\text{real}*}) > NE(e^{\text{fix}})$, this foreign demand exceeds the private supply coming from exports, putting downward pressure on the DOM-FOR FX rate.

Central Bank Reserves

If the domestic central bank is not able to meet the difference between desired capital outflows $CF(r_d^{\text{real}*})$ and net exports $NE(e^{\text{fix}})$ with its reserves of the foreign currency, then either (i) the FX peg must break, or (ii) the interest rate must revert to the higher level, r_d^{real}, or a combination of these factors. In case (i), the domestic currency weakens to e^* such that $NE(e^*) = CF(r_d^{\text{real}*})$ and equilibrium is restored. In case (ii) the private investors buy FOR currency from the central bank's stock of FOR reserves, and sell it the DOM currency. The central bank's reabsorption of the DOM currency may undo its initial monetary stimulus by shifting the LM curve back left, leading to the interest rate reverting to r_d^{real}.

If the domestic central bank has ample foreign FX reserves, then it may be able to "bridge" the gap between $CF(r_d^{\text{real}*})$ and $NE(e^{\text{fix}})$ for a period of time. During this period, it effectively satisfies some or all of the private demand for foreign deposits (and investments) and therefore shifts the CF curve to the left, allowing the real FX rate to remain at e^{fix}. It may do this as long as its reserves last.

The Euro and the Impossible Trinity

A closely related idea to a fixed FX rate is that of a common currency. For example, since the implementation of the monetary union in 1999, and the introduction of physical Euros in 2002, the countries that have adopted the Euro have effectively fixed their exchange rate to each other. The bloc of 19 countries (at the time of writing) that use the Euro is called the Eurozone. Further, since the Eurozone countries allow free capital flows between them, it is clear that they have chosen to be on side B of the triangle in Fig. 6.6 and therefore cannot have monetary autonomy. Indeed, for this reason the single currency is managed by a single monetary authority, the ECB, as predicted by the Impossible Trinity theory.

The next feature box discusses how the fixed FX rate and loss of monetary autonomy exacerbated the Eurozone debt crisis that began in the aftermath of the GFC.

While the Impossible Trinity theory provided the restrictions on either capital mobility or monetary autonomy that are associated with a fixed FX rate, it did not examine economic implications of such a policy. This is the topic of the next section.

6.4 Optimal Currency Area (OCA)

Few economists today argue that the world should run a system of fixed exchange rates (such as Bretton Woods), or that there should be only one single common currency. At the opposite end of the spectrum, few economists argue that each state in the U.S. or each city in the U.K. should issue its own currency. How, then, should we decide the scope of geographic regions that share a currency? Mundell (1961) theorised the criteria in which a currency union is beneficial, calling it an *Optimal Currency Area* (OCA).

In short, he argues that at least the following four criteria are required: (i) high labour mobility across the area, (ii) capital mobility, and price and wage flexibility, (iii) a fiscal risk sharing mechanism and (iv) closely related business cycles. If these basic conditions are met, then the area may benefit from a shared currency, and it can be called an OCA.

The benefits of an OCA are that areas that share a currency face lower transactions costs when trading with each other, and eliminate FX volatility risk, potentially leading to increased trade. Further, a central monetary authority may scale itself to competently manage the monetary affairs of a wider region. Finally, there may be additional political and social benefits to increased economic integration, such as a decrease in the probability of conflict and war. Indeed, this is likely to have been a major motivation behind the Euro.

However, the feature box discusses how the Eurozone is not a natural OCA compared with, say, the states of the U.S., and how the consequences of this played out during the Eurozone debt crisis.

The Euro, the Eurozone Crisis, Ireland and Iceland.

While the U.S. and much of the rest of the world began recovering from the GFC in late 2009, this period marked the beginning of second crisis in the Eurozone. The Eurozone claim to being an OCA had weaknesses. For example, compared to the U.S. there is low labour mobility across Europe, perhaps due to language and cultural barriers across countries. There was also no mechanism for fiscal transfers. If unemployment in a U.S. state rises, then benefits are paid to unemployed people in that state, paid for from the aggregate taxes collected by the U.S. government from all states. No such mechanism exists in the Eurozone. This lead to a need for the countries experiencing weak post GFC growth, such as Portugal, Spain, Italy, Greece and Ireland, to run more stimulative monetary policy (through the mechanism described in Sect. 6.2.4) than others, such as Germany and France.

The single currency and the associated inability for any individual Eurozone member to devalue its FX rate and run stimulative monetary policy have been cited as exacerbating factors for the crisis (see, for example, Lane, 2012). Since the EUR FX rate did not adjust to the degree required for the most indebted nations, the only remaining mechanism for these economies to become more competitive was an internal price adjustment—a deflation. This deflation added to debt overhang problems in already indebted nations.

The remainder of this feature box provides a case study to illustrate these important ideas by comparing the outcomes of Ireland, a Eurozone member, and Iceland, which has an independent currency, through this period.

Iceland and Ireland

Both Ireland and Iceland entered the GFC in a similar position to each other. The banking sector in both of these countries had quickly scaled up their balance sheets since 2000, and particularly quickly during the GSG period from 2003 to 2007. In Iceland, total bank assets rose from 200% to 1000% of GDP over the GSG period. In Ireland, the total value lending to government, corporates and households had risen from approximately 100% to 300% of GDP over this period. For comparison, over the same period, in the Euro Area as a whole, this rise was from approximately 150% to 200% of GDP, and in the U.K. from 150% to 250% of GDP.[6]

As the crisis broke, and funding markets no longer allowed the banks of either country to fund their loan books, Iceland and Ireland's prospects diverged. The Irish government announced that it would guarantee all liabilities for seven of its banks to a total of 375 billion EUR, approximately 2 times GDP. However, it was well understood in the markets that this was too much of a burden for the government to bear, and this would lead to a debt crisis. Meanwhile, Iceland announced that it would make its domestic depositors whole, but default on its obligations to international investors.[7] However, at this point the outcomes of the two countries begin to diverge.

Figure 6.8 plots the FX rate of the EUR and the Icelandic Krone (IKR). The IKR devalued between 2008 and 2010, and did not recover. Meanwhile, the EUR oscillated wildly, falling at times, but then recovering ground. Had the Euro not been jointly backed by the stronger sovereigns of Germany, France and others, then it too could have devalued. This represents a failure to fulfil the criteria of convergent business cycles, required for an OCA.

[6] Source "Misjudging Risk: Causes of Systemic Banking Crisis in Ireland," Report of the Commission of Investigation into the Banking Sector in Ireland.

[7] A large number of deposits had been taken from U.K. households. The U.K. Treasury eventually made households whole, but expressed its displeasure in bearing this cost by at one point placing an Icelandic bank, Landsbanki, on its watch list of rogue organisations that included terrorist groups.

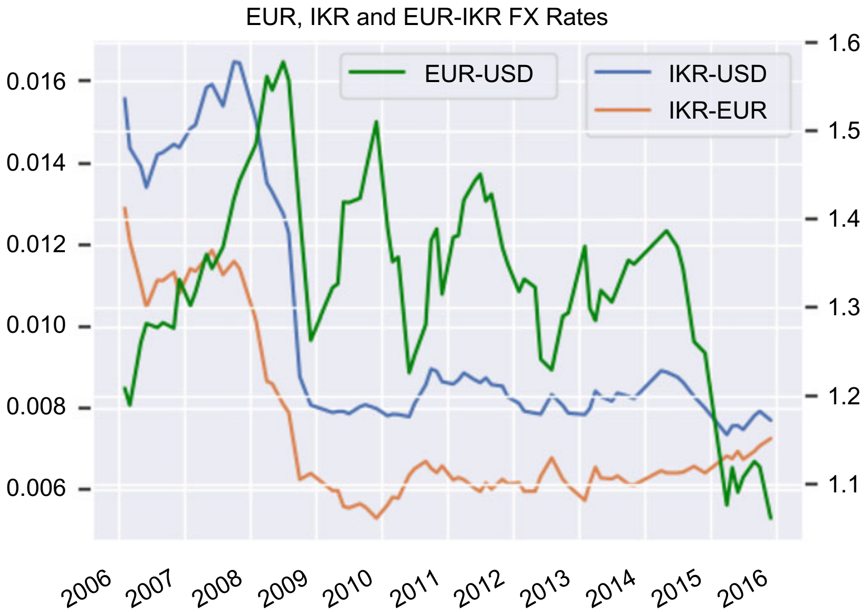


Fig. 6.8 The figure shows the price in USD of the EUR (right-hand axis), and the price of the IKR in both USD and EUR. The IKR significantly devalued against both the EUR and the USD in 2009, and did not rebound. The EUR fluctuated over this period. The devaluation in the IKR may have helped Iceland's recovery. Board of Governors of the Federal Reserve System (US), Foreign Exchange Rates [DEXUSEU, CCUSMA02ISM618N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DEXUSEU, October 22, 2021

Next, Fig. 6.9 shows Iceland and Ireland's divergent inflation paths. Both had inflation in the region of 5% going into the crisis. However, these paths move in opposite directions. While Iceland allowed inflation to rise to 15%, Ireland experienced a deflation. Inflation in Ireland ran negative again in 2014 as the crisis continued. The deflation added to Ireland's debt overhang problems.

Finally, Fig. 6.10 shows real GDP. Real GDP in Ireland was a little under that in Iceland at the depths of the GFC in 2010. However, since that point, the figure shows that Iceland performed significantly better. It was not until 2015 that Ireland was able to catch up. However, note that this catch up also corresponded to a fall in the EUR.

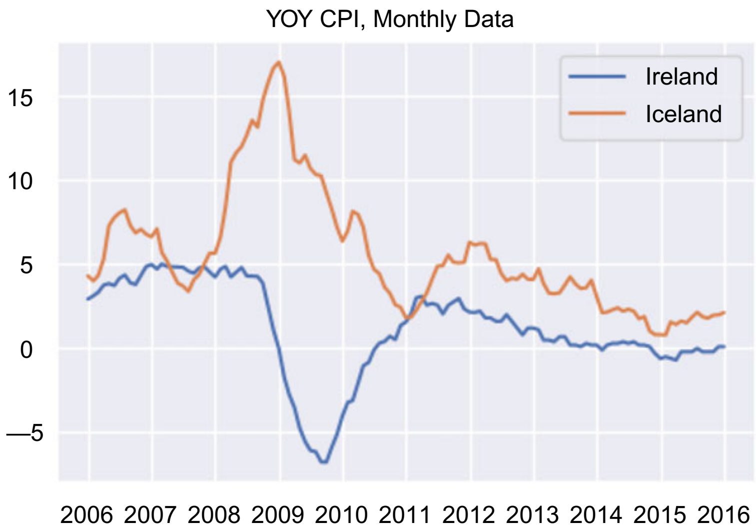


Fig. 6.9 The figure shows YOY CPI, using monthly data in Ireland and Iceland. Both countries enter the crisis at similar levels, and despite similar macroeconomic situations, they experience significantly different inflation paths. High inflation in Iceland decreased the real value of the debt, whereas deflation in Ireland caused an increase (*Source* Board of Governors of the Federal Reserve System [US], CPI Data [IRLCPIALLMINMEI, ISLCPIALLMINMEI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DEXUSEU, October 22, 2021)

Putting these charts together, one may argue that allowing a devaluation in the currency and high inflation is able to set a country on a path to real GDP growth faster than otherwise. While the single currency may have many benefits, the further away the system is from being an OCA, it may therefore have significant costs also.

6.5 Chapter Summary

- An increase in the money supply boosts output and weakens the domestic currency. However, it also lowers interest rates and increases net exports and capital outflows.
- The extended Mundell-Fleming model tells us that fiscal stimulus boosts output and strengthens the domestic currency. Simultaneously, interest rates rise, and net exports and capital outflow decrease.

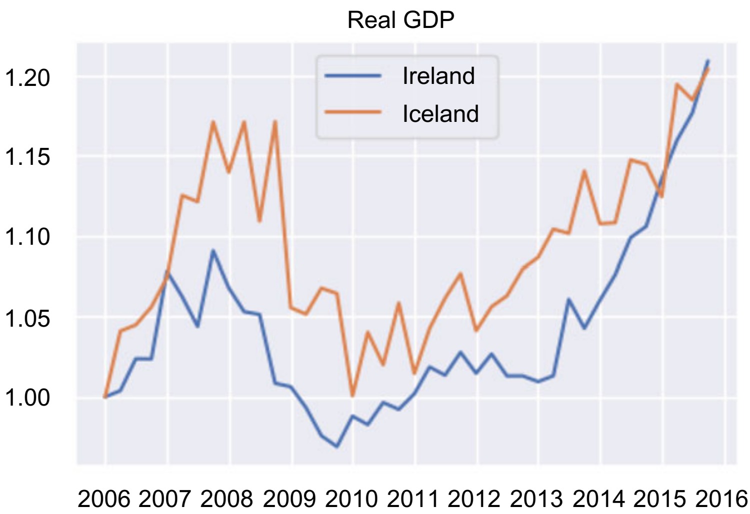


Fig. 6.10 The figure plots real GDP growth in Ireland and Iceland. While both real GDP plunged in 2010, Iceland recovered more quickly. This may have been due to higher inflation reducing the real value of its debt, and a weaker FX rate increasing its competitiveness. Note that Ireland's eventual catch up coincides with the fall in the EUR FX rate (*Source* Board of Governors of the Federal Reserve System [US], Real GDP Data [CLVMNACSAB1GQIE, CLVMNACSAB1GQIS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DEXUSEU, October 22, 2021)

- However, the Mundell-Fleming approach does not account for changes in expectations about future taxes. If a fiscal stimulus is unaffordable, then consumers may save more to offset the stimulus. This is called Ricardian equivalence. Alternatively, if consumers expect the government to roll over the additional debt incurred to fund the stimulus, then they may expect a future monetary expansion, a shift left in the LM curve, inflation and thereby a weakening in the FX rate.
- The Impossible Trinity expresses the idea that a country cannot have all three of (i) free capital flows, (ii) a fixed FX rate and (iii) monetary autonomy. The idea can be proven via the CIP relationship when there exist default free deposits in both the foreign and domestic economy. It can also be proven more generally via the extended Mundell-Fleming model.
- A common currency is similar to a fixed FX rate. There are benefits and costs to sharing/fixing a currency. The case study comparing the Ireland and Iceland's outcomes post the GFC illustrates the potential costs that can occur if the common currency is over a non-OCA area.

References

Hicks, J. R. (1937). Mr. Keynes and the classics: A suggested interpretation. *Econometrica, 5*, 147–159.

Keynes, J. M. (1936). *The general theory of employment: Interest and money*. Macmillan.

Lane, P. R. (2012). The European Sovereign debt crisis. *Journal of Economic Perspectives, 26*(3), 49–68.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy, 1*, 19–46.

Mankiw, N. (2007). *Macroeconomics* (6th ed.). Worth Publishers.

Mundell, R. A. (1961). A theory of optimum currency areas. *The American Economic Review, 51*(4), 657–665.

7

Inflation

Inflation and inflation expectations play an important role in FX determination. We studied why in Chapter 3. To recapitulate, the idea was that in fixed real FX rate models, higher expected inflation decreases the nominal forward value of a currency to keep its expected real purchasing power unchanged. Whether the spot FX rate strengthens or weakens then depends on the strength of the reaction of the central bank. The currency strengthens if it is larger than the final increase in inflation expectations, plus the change in the risk premium, as described in Sect. 3.5. FX investors must therefore study inflation across countries and form a view on its dynamics as an input into FX price modelling.

A detailed description of the theories of inflation is beyond the scope of this chapter. Indeed, the topic of monetary economics is itself broad enough to fill several textbooks. I therefore provide a survey of, and introduction to, the main ideas in the area to a level that facilitates a practical understanding of the key variables that are referred to by central bankers, policy markers, academic economists and others.

First, I discuss the main inflation measures, namely the consumer prices index (CPI), the GDP deflator and the personal consumption and expenditures price index (PCE). I focus on the U.S., but analogous metrics are used across world economies.

Second, I survey the main classical theories in monetary economics. I begin with the Aggregate Demand (AD)-Aggregate Supply (AS)/Phillips curve model that is built on Keynesian/IS-LM foundations, before moving

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9_7

on to monetarism and the quantity theory of money commonly associated with economist, Milton Friedman.

Finally, I provide a short description of more recent thinking in monetary economics, namely the Fiscal Theory of the Price level (FTPL), which argues that inflation is determined by the tax revenues that back government liabilities, and the so-called new-Keynesian Phillips curve (NKPC) models that permeate much central bank and academic thinking today.

With the main theories of inflation in place, the reader will understand the models that are needed to frame important questions that FX investors face on a regular basis such as, and among others, when is a fiscal stimulus inflationary? Is QE inflationary? Is helicopter money inflationary? Why do unsustainable government debt dynamics create inflation? Such understanding provides a better ability to forecast interest rates and form inflation expectations, and thereby determine FX rates through the methods that we have studied.

7.1 Measuring Inflation

The reason that there are so many methods of measuring inflation is that the task is fraught with difficulties that stem from the fact that the prices of multiple real goods must be included. In the hypothetical single good *widget* economy that we studied in Chapter 3, inflation is straight forward to calculate; simply track the price of the single item over time. However, when multiple goods are included, measurement biases are all but inevitable.

The CPI is often thought to overstate inflation because it is typically a Laspeyres index, meaning that it is calculated by tracking the prices of a fixed basket of goods. If, for example, the prices of oranges rise, consumers may substitute away from oranges towards, say, apples, and thereby avoid some of the price increase. The fixed basket in the CPI calculation does not account for this and therefore overweights oranges relative to the actual consumer basket.

The GDP deflator is a Paasche index, meaning that it is calculated by tracking the prices of a basket of goods that changes with consumer behaviour, and therefore captures the impact of consumers substituting cheaper goods for more expensive ones. However, this measure may understate inflation. Presumably consumers draw less utility from switching away from their preferred fruit! If inflation is thought of in a pure sense as the price of a unit of utility then changing the basket Paasche methods does not capture this.

PCE offers a middle ground. It is calculated using the Fisher Index, which is the square root of the product of a Laspeyres index and a Paasche index. I discuss these ideas in more detail over the remainder of this section.

Finally, note that it is the same difficulty that makes measuring inflation challenging, namely incorporating multiple goods, that makes FX rate determination difficult. In the single good economy, the PPP approach to the FX rate rests on strong ground, and more so if the foreign and domestic widgets were comparable in quality and easily transportable. Incorporating multiple goods makes the real FX rate difficult to determine because it is not clear which foreign and domestic basket should be compared. Further difficulty arises from assessing the relative and changing quality of foreign and domestic goods and service, and also from the different inflation baskets used to calculate inflation metrics across nations.

To understand these challenges in more detail, let us now discuss the inflation metrics in more detail.

7.1.1 CPI

The most commonly used measure of the general price level in an economy is the CPI. The CPI is used as the reference inflation rate for financial contracts such as Treasure Inflation Protected Securities (TIPS) and inflation swaps. It is also used to adjust social security payments, among other uses.

The CPI is a basket that attempts to replicate the consumption of the average household. The full calculation of U.S. CPI is available from the website of the Bureau for Labor Statistics (BLS), but it is complicated and beyond the scope of this chapter. Here, I provide a simplified version to provide the important intuition behind the measure.

The CPI index for month T and for a base year b is given by,

$$CPI_{b,T} = \prod_{t=b}^{T} \Delta CPI_t,$$

where

$$\Delta CPI_t = \frac{\sum_{i=1}^{n_b} q_y^i p_t^i}{\sum_{i=1}^{n_b} q_y^i p_{t-1}^i} \text{ for } t > b, \tag{7.1}$$

and

$$\Delta CPI_t = 100 \text{ for } t = b. \tag{7.2}$$

q_y^i is the quantity of the i-th good bought by the consumer based on the most recent expenditure survey, assumed to be carried out at time point y. p_t^i is the price of the i-th good at time t, and n_b is the number of items in the basket.

In words, Eq. (7.2) says that the CPI index is normalised to 100 in the base year, and Eq. (7.1) says that ΔCPI_t is the relative change in the CPI index from month $t - 1$ to t. In the U.S. the CPI index typically reported in the media uses 1982–84 as its base year.[1]

The quantities in the CPI basket do not change until the basket is updated. In the U.S., this is typically every two years. That is, q_y changes every two years. In between basket updates, it is therefore only prices that can move CPI_t higher or lower. CPI therefore does not capture changes in changes in the consumer basket from month to month, but does so with a long time lag. This may lead to an overstatement of inflation, as consumers substitute away from more expensive goods and towards cheaper ones.

Relative Importance

In the U.S., food and energy constitute approximately 15% and 6% of the basket, respectively. Although these items together contribute just 21% of the total basket, their highly volatile price series have lead policymakers and market participants to separately monitor the so-called *core* CPI, which excludes these components so as to provide observers with a clearer measure of inflationary trends.

Table 7.1 shows relative importance in the U.S. CPI. These weights naturally vary across countries, with the main trend being that food forms a greater part of the basket in emerging market nations than in developed nations.

CPI Growth Rates

Media reports may cite CPI in several different forms. Below, I list examples in three common formats.

The Year-Over-Year (YOY) CPI from Q2 2020 to Q2 2021 is

$$\text{YOY} = \frac{CPI_{Q2-21}}{CPI_{Q2-20}} - 1. \tag{7.3}$$

The Quarter-Over-Quarter (QOQ) CPI from Q1 2021 to Q2 2021 is,

$$\text{QOQ} = \frac{CPI_{Q2-21}}{CPI_{Q1-21}} - 1. \tag{7.4}$$

[1] The BLS has normalised so that the average index level for the 36-month period covering the years 1982, 1983 and 1984 is equal to 100 and then measures changes relative to that point.

Table 7.1 The table shows the relative importance of the different category groups in the U.S. CPI basket in December 2020. The relative importance of a component is its value weight expressed as a percentage of all items. In addition to these broad categories, there are several commonly cited subcomponents and combinations of subcomponents that market participants address. For example, *energy* (6.2%) is the sum of *motor fuel* (2.9%), which is a subcomponent of transport, and *household energy* (3.3%), which is a subcomponent of housing. *Shelter* (33%) is commonly cited subcomponent of housing. Transportation services (5.1%) are transport (15.2%), less new and used vehicles, excluding those rented or leased (6.8%), motor fuel (2.9%) and motor vehicle parts and equipment (0.4%)

Group	Relative Importance
Food and Beverages	15.5
Housing	42.4
Apparel	2.7
Transport	15.2
Medical care	8.9
Recreation	5.8
Education	6.8
Other goods	3.2
Total	100

Source Website of the U.S. Bureau of Labor Statistics

The QOQ CPI from Q1 2021 to Q2 2021 is,

$$\text{QOQ Annualised} = \left(1 + \frac{CPI_{Q2-21}}{CPI_{Q1-21}}\right)^4 - 1. \tag{7.5}$$

Similar equations hold for the GDP deflator and the PCE.

7.1.2 GDP Deflator

The GDP deflator is the ratio of nominal GDP to real GDP. It is given by

$$\text{GDP Deflator}_t = \frac{\text{Nominal GDP}_t}{\text{Real GDP}_y} = \frac{\sum_{i=1}^{n_a} q_t^i p_t^i}{\sum_{i=1}^{n_a} q_t^i p_y^i} \times 100. \tag{7.6}$$

The y subscript denotes the base year relative to which the variables are calculated, q_t^i and p_t^i are the quantity produced and price of the i-th good in year t, respectively, and n_a is the number of goods and services produced. Like CPI, the GDP deflator reflects how overall prices in the economy are moving. However, there are three main differences.

First, the GDP deflator measures the prices of all n_a goods and services produced, whereas CPI measured only the n_b items in the consumer basket.

Therefore, goods and services purchased by firms and the government will impact the GDP deflator, but get ignored by CPI.

Second, the GDP deflator reflects domestic production. A rise in the price of imported goods impacts CPI, but not the GDP deflator. Note that this difference is important for real FX rate calculations.

Third, as previously discussed, the GDP deflator is a Paasche index, whereas CPI is a Laspeyres index. While CPI may therefore overstate inflation, the GDP deflator may understate it.

7.1.3 PCE

Although CPI is more commonly cited in media reports, the FOMC statement of January 2012 noted that the "personal consumption expenditures, is most consistent over the longer run with the Federal Reserve's statutory mandate," thereby suggesting that it is the Federal Reserve's preferred inflation metric.

There are several important differences between the PCE and CPI. The first is that PCE uses a Fisher index. If the Laspeyres index (used to calculate CPI) is denoted by L, and the Paasche index (used to calculate the GDP deflator) is P, then the reader may think of the Fisher index as $\sqrt{LP}$. It is therefore a middle ground between L and P. The full PCE calculation is beyond the scope of this chapter. However, the important takeaway is that by incorporating P, PCE is able to account for the substitution effect of consumers purchasing cheaper goods, whereas the CPI is not. This is likely a major contributor towards PCE reporting lower inflation than CPI (see Fig. 7.1).

Second, CPI and PCE are based on different surveys. The CPI is based on a survey of what households are buying; the PCE is based on surveys of what businesses are selling. The third is that CPI does not include indirect expenses (such as healthcare paid by an employer, for instance), whereas these are included in PCE.

7.2 AD-AS Models

The AD-AS model is built on top of the IS-LM approach of the previous chapter, coupled with a so-called expectations-adjusted Phillips curve (more on this later). It is the most common model used to understand inflation in that it is the model that is implicit in statements made by the U.S. Federal Reserve, media reports, and by many economists writing for non-academic

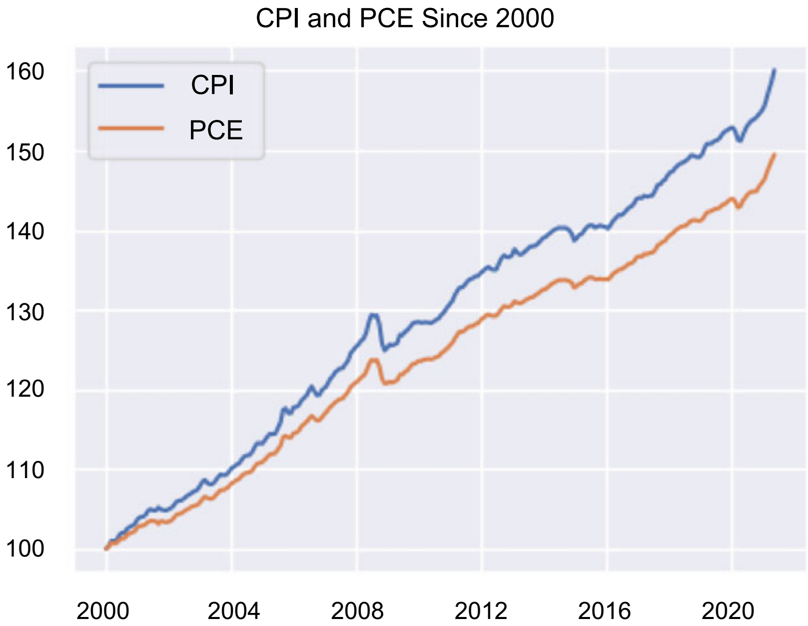


Fig. 7.1 The figure compares the CPI and PCE indices, normalised to 100 in January 2000. The CPI index has grown at a faster rate than PCE. The data shows that CPI averaged 2.22% gross growth per year, compared with PCE at 1.9%. The CPI-PCE wedge was 0.32%. Core CPI and core PCE are not plotted here. However, the average growth rates of these quantities were 2.07% and 1.8% over the period (*Source* U.S. Bureau of Economic Analysis, Inflation Data [CPIAUCSL, CPILFESL, PCEPI, PCEPILFE], retrieved from FRED, Federal Reserve Bank of St. Louis, August 19, 2021)

professional audiences. It is therefore an important model for investors to understand. However, the reader may be left unsatisfied, because the model does not contain features such as utility functions and thereby risk premiums, budget constraints, time consistent investment-consumption decisions or other features one may expect from a modern economic model. It therefore does not contain the central pricing equation of Chapter 1 that has been key to the understanding of FX markets built over this book. Nevertheless, it provides useful intuition, and its widespread use makes it important for us to study here. Later in the chapter, I present models that are better able to withstand the Lucas critique.

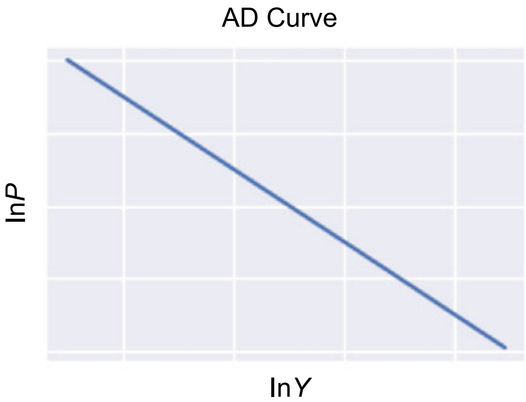


Fig. 7.2 The AD curve plots the relationship between the price level in the economy P and the level of output Y, such that the goods market (IS curve) and money markets (LM curve) are both clear. If P is low, Y is high. For convenience, I plot $\ln P$ against $\ln Y$ and draw a straight line, although the curve may take other shapes

7.2.1 The AD Curve

The AD curve plots the relationship between the price level P and the quantity of goods and services demanded, Y. Its downward slope is intuitive because it implies that there is less demand at higher prices and greater demand at lower prices. It is shown in Fig. 7.2.

The derivation of the AD curve requires the IS and LM curves of Sect. 6.2. I proceed with an intuitive derivation before offering an (optional) formal version.

We wish to understand how Y reacts to a change in P. Suppose that P increases. This leads to a decline in real money supply because M/P is lower. For the money market to clear, real money demand, $L(r_d, Y)$, must therefore fall (recall Eq. [6.2] equated real money supply with real money demand). Since $L(r_d, Y)$ depends on r_d and Y, this can only happen through a change in Y, r_d, or both. However, the IS curve already provided a restriction on how r_d and Y can move to clear the goods market. Its downward slope means that if r_d rises then Y must fall (recall Fig. 6.8). There we have our answer. The only way for $L(r_d, Y)$ to fall is if Y falls and r_d rises, because $L(r_d, Y)$ depends positively on Y and negatively on r_d (see Sect. 6.1.2).

Put differently, the IS curve tells us that if P were to rise and Y were to rise, then r_d would fall, and this would lead to $L(r_d, Y)$ rising, pushing the money market out of equilibrium. The AD curve is therefore downward sloping.

Deriving the AD Curve (Optional)
First, calculate the derivate of both sides of Eq. (6.2),

$$\mathrm{d}L = \frac{\partial L}{\partial r_d}\mathrm{d}r_d + \frac{\partial L}{\partial Y}\mathrm{d}Y = \mathrm{d}\left(\frac{M}{P}\right). \tag{7.7}$$

The right-hand side is the change in the real money supply that results from a change in the price level P. It is negative when P rises. The left-hand side is the change in L. The first equality says that the change in L must result from changes in r_d and Y.

Second, recall that the IS curve imposes that equilibrium in the goods market requires that $\frac{\mathrm{d}Y}{\mathrm{d}r_d} < 0$. Therefore, we may write $\mathrm{d}Y = -\alpha \mathrm{d}r_d$, where $\alpha > 0$. Substituting this into Eq. (7.7) and rearranging, we have,

$$\frac{\mathrm{d}Y}{\mathrm{d}\left(\frac{M}{P}\right)} = \frac{1}{-\frac{1}{\alpha}\frac{\partial L}{\partial r_d} + \frac{\partial L}{\partial Y}}. \tag{7.8}$$

Finally, noting that

$$\mathrm{d}\left(\frac{M}{P}\right) = -\frac{M}{P^2}\mathrm{d}P, \tag{7.9}$$

we have

$$\frac{\mathrm{d}Y}{\mathrm{d}P} = -\frac{M}{P^2}\frac{1}{-\frac{1}{\alpha}\frac{\partial L}{\partial r_d} + \frac{\partial L}{\partial Y}} < 0. \tag{7.10}$$

Since $\frac{\partial L}{\partial r_d}$ is negative, and $\frac{\partial L}{\partial Y}$, M, P and α are all positive, the right-hand side of Eq. (7.10) is negative. Therefore, an increase in P leads to a decrease in Y, and the AD curve is downward sloping as required.

7.2.2 The AS Curve

The AS curve plots the relationship between the price level and the quantity of goods supplied, Y. I show in Sect. 7.3 that the AS curve and the Phillips curve are two sides of the same coin. The AS curve is typically written as,

$$\ln Y = \ln \bar{Y} + \alpha(\ln P - \mathbb{E}[\ln P]). \tag{7.11}$$

Here, $\bar{Y}$ denotes the natural level of output, and $\mathbb{E}[\ln P]$ is the expected (log) price level. This equation is intuitive. It says that firms supply more goods and services if prices turn out to be higher than expected, and fewer if prices turn out to be lower. Therefore, at higher prices, output Y is greater than its natural level $\bar{Y}$. Figure 7.3 plots the AS curve.

There are three main theories used to derive this relationship, namely the *sticky-price model*, the *sticky-wage model* and the *imperfect information model*. Here, I ask the reader to take Eq. (7.11) as given, so that we may begin applying the model to understand its implications for inflation, and refer the reader to standard texts in macroeconomics, such as Mankiw (2007) for a detailed analysis of these theories.

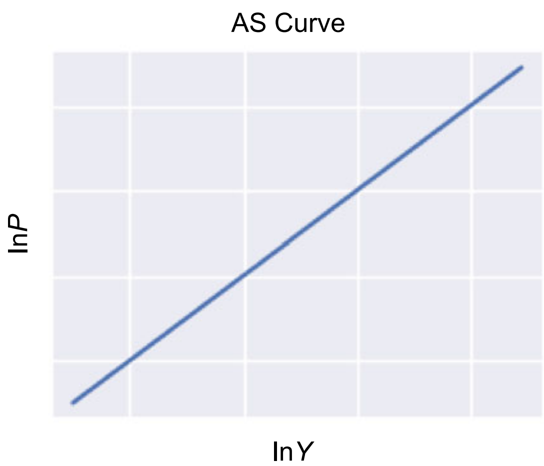


Fig. 7.3 The AS curve shows the relationship between the price level in the economy P and the goods supplied Y. It is upward sloping implying that output Y is higher when prices are high

7.2.3 Demand-Pull Inflation

Suppose that the central bank aims to stimulate the economy. It lowers the interest rate through an increase in the money supply. This shifts the AD curve to the right. The reason is that, for a given level of P, an increase in M raises the real money supply, M/P. For the money market to clear, real money demand, $L(r_d, Y)$ must therefore increase to match the increased supply. The IS curve restricts Y and r_d to move in opposite directions (as discussed in Sect. 6.2). If Y increases and r_d decreases, then both act to increase $L(r_d, Y)$. Therefore, a higher level of M requires a higher level of Y for any given level of P. If Y is higher at every level of P, the AD curve has shifted to the right. Figure 7.4 shows this shift, and also how the economy evolves in response.

Let us assume that we start at point 1, where output Y is equal to potential output $\bar{Y}$, indicated by the vertical dotted line. The monetary expansion shifts the AD curve to the right. However, if the monetary expansion was unanticipated, the model assumes that the AS curve remains unchanged in the short run because $\mathbb{E}[\ln P]$ had already been set. This shifts the economy to point 2, where both output is higher and prices are higher. That is, the economy experiences a boom. Over time, this boom leads to inflationary pressures, as unemployment falls below its natural rate, and wages rise. This finally causes an increase in $\mathbb{E}[\ln P]$, leading to a shift higher in the AS curve. The economy then contracts back to point 3, where output is back to its potential level $\bar{Y}$, but prices are higher.

7.2.3.1 Implications for the FX Rate

The fixed real FX rate approach of Chapter 3 tells us that nominal forward price of the DOM currency should be lower in response to the rising price level. This ensures that the FOR currency is able to purchase more of the DOM currency to offset the rising prices, and maintain the real FX rate unchanged. Further, the CIP relationship of Chapter 2 implies that the spot FX rate weakens even more so than the forward FX rate due to the lower nominal interest rate that results from the increase in M.

A weakening in the DOM currency is also consistent with the extended Mundell-Fleming approach of Chapter 6. The increase in M lowers the interest rate, causing an increase in capital outflows and a corresponding increase in net exports.

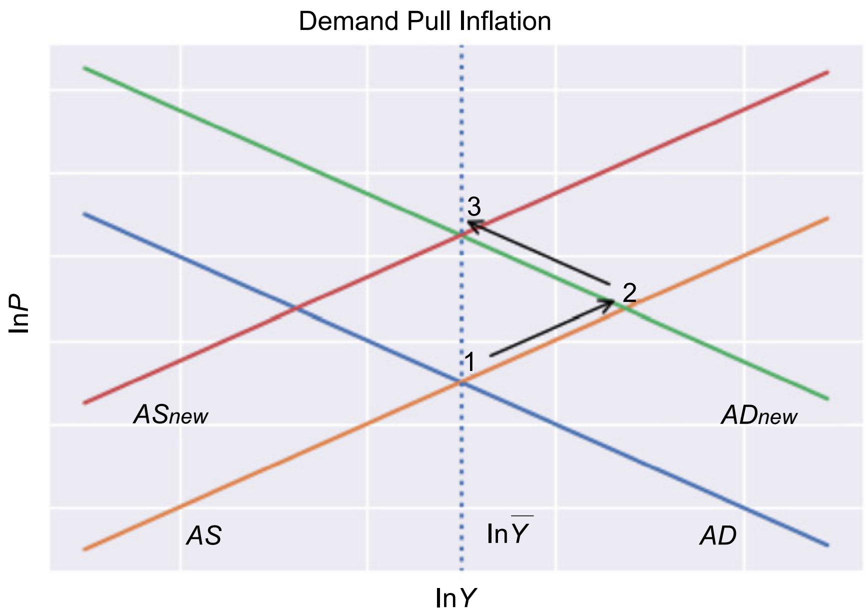


Fig. 7.4 The AD curve shifts to the right in response to a monetary expansion designed to lower the interest rate from the *AD* curve to the AD_{new} curve. This moves the economy from point 1 to point 2, where both output and prices are higher. As participants revise higher their price expectations, and thereby raise $\mathbb{E}[\ln P]$, the AS curve shifts higher from *AS* to AS_{new}. This causes the economy to move to point 3. Output returns to its potential level $\bar{Y}$, but prices reset at a higher level

7.2.4 Cost-Push Inflation

Suppose that there is a supply shock, such as an increase in the prices of commodities, or raw materials. This can cause the AS curve to shift to the left, because the higher prices of inputs cause firms to produce less for any given price level.

Figure 7.5 shows how the economy may respond. Let us assume that we start at point 1, where output Y is equal to potential output $\bar{Y}$, indicated by the vertical dotted line. The left shift in the AS curve causes the economy to move to point 2, where output is lower, but prices are higher. That is the economy experiences a stagflation, with unemployment likely to increase.

The figure shows how the economy reacts if the central bank lowers the interest rate, or the government increases G. Both approaches shift the AD curve to the right, thus moving the economy back to point 3. Output returns to its potential level, but at a yet higher price level.

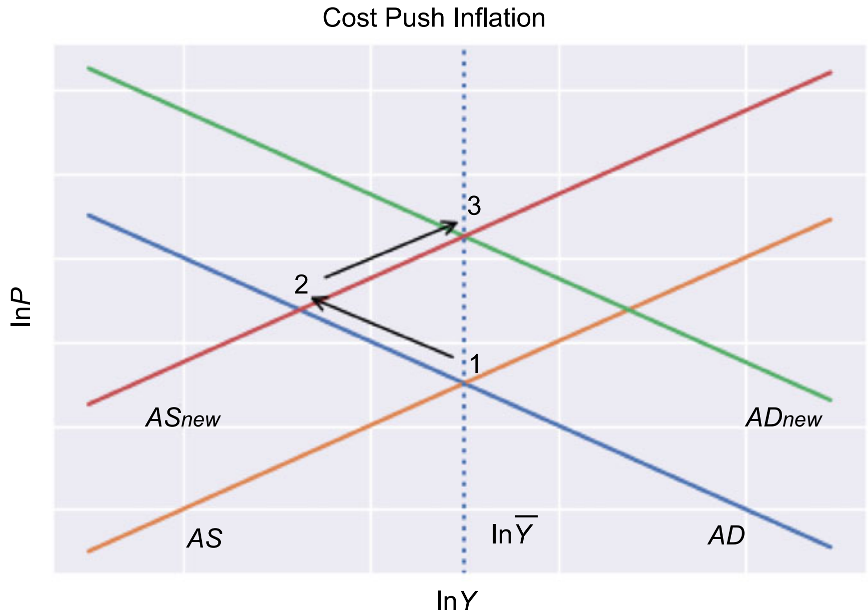


Fig. 7.5 The AS curve shifts higher from *AS* to AS_{new} in response to a supply shock. This moves the economy from point 1 to point 2, where prices are higher, but output is lower, causing unemployment to rise. The government and/or central bank may respond to the shortfall in employment by acting to shift the AD curve from *AD* to AD_{new}. This moves the economy back to point 4, where output is at its potential level $\bar{Y}$, but prices reset at a higher level

This example shows the importance in understanding the source of inflation. A central bank with a mandate focused solely on realised inflation may hike interest rates in response to the economy moving from position 1 to position 2, because the price level has risen. Let us consider the implications of such a move.

The hike in the interest rate causes the AD curve to the left. This will indeed cause the price level to fall, but does so at the cost of even lower output and therefore the precipitation of the economic turmoil. The feature box provides an example of such a circumstance.

Cost-Push Inflation and the Central Bank Response

In 2011, during the European debt crisis, the European Central Bank (ECB), lead by Jean-Claude Trichet, famously hiked the ECB Deposit Facility Rate from 0.25% to 0.75% in two 0.25% increments in response to rising inflation. The statement that accompanied the hikes said "upward pressure on inflation, mainly from energy and commodity prices," thereby citing some cost-push factors.

As the crisis deepened, output continued to fall. The hike was subsequently repealed by Trichet's successor, Mario Draghi, later in the year. Inflation continued to fall with and the ECB continued to attempt to shift the AD curve to the right by embarking on further *dovish* path that lead to the negative interest rate that remains at the time of writing.

Implications for the FX Rate

Just as in the case of demand-pull inflation, the fixed real FX rate of Chapter 3 tells us that nominal forward price of the DOM currency should be lower in response to the rising price level. This ensures that the FOR currency is able to purchase more of the DOM currency to offset the rising prices, and maintain the real FX rate unchanged. However, the key difference is that with cost-push inflation, the price level rises and the FX rate weakens *before* the interest rate changes, as the economy moves from point 1 to point 2.

7.3 The Phillips Curve

The Federal Reserve operates under a mandate from Congress to "promote effectively the goals of maximum employment, stable prices, and moderate long term interest rates." Central banks around the world generally share similar mandates.

Through the AD-AS model, we understood that there is a trade-off between real GDP growth and inflation. To recapitulate, the central bank may raise M to shift the AD curve to the right, and thereby boost output, but in doing so raises the price level. The Phillips curve[2] re-expresses this idea as a trade-off between unemployment and inflation. In doing so, we are able

[2] The Phillips curve is named after New Zealand economist A. W. Phillips after his plot published in 1958 of the negative relationship between inflation and unemployment in the U.K. over the period 1861 to 1957.

to understand why the central bank's mandate is written in this form, and better understand the actions that a central bank may implement.

Let us derive the Phillips curve. First, write the AS curve as,

$$\ln P_t = \frac{1}{\alpha}(\ln Y_t - \ln \bar{Y}) + \mathbb{E}_{t-1}[\ln P_t] + \varepsilon_t. \tag{7.12}$$

This is simply a re-arrangement of Eq. (7.11), with the addition of the term ε_t. I add this term to denote a random supply shock that may raise or lower prices, such as that discussed in Sect. 7.2.4. Next, subtract $\ln P_{t-1}$ from both sides of Eq. (7.12),

$$\ln P_t - \ln P_{t-1} = \frac{1}{\alpha}(\ln Y_t - \ln \bar{Y}) + \mathbb{E}_{t-1}[\ln P_t - \ln P_{t-1}] + \varepsilon_t. \tag{7.13}$$

Letting $\pi_t^{\text{act}} = \ln P_t - \ln P_{t-1}$ denote the actual rate of inflation from time $t-1$ to time t, and $\pi_{t-1}^e = \mathbb{E}_{t-1}[\ln P_t - \ln P_{t-1}]$ denote the expected rate of inflation from time $t-1$ to time t, calculated at time $t-1$, we may rewrite Eq. (7.13) as,

$$\pi_t^{\text{act}} = \pi_{t-1}^e + \frac{1}{\alpha}(\ln Y_t - \ln \bar{Y}) + \varepsilon_t. \tag{7.14}$$

The final step is to add in the relationship between the output gap and unemployment,

$$\left(\frac{1}{\alpha}\right) \ln Y_t - \ln \bar{Y} = -\beta(u_t - \bar{u}). \tag{7.15}$$

Here, u is the unemployment rate and $\bar{u}$ is the natural rate of unemployment. This is a version of Okun's law, named after economist Arthur Okun. It states that the output gap is proportional to the deviation in unemployment from its natural level. Substituting this into Eq. 7.14 give us,

$$\pi_t^{\text{act}} = \pi^e - \beta(u_t - \bar{u}) + \varepsilon_t. \tag{7.16}$$

Equation (7.16) is the Phillips curve. In words, it states the rate of inflation depends on (i) expected inflation, (ii) the deviation of the unemployment rate from its natural level, or *cyclical unemployment*, and supply shocks.

Written in this form, the Phillips curve makes clear the policy maker's dilemma; a mandate of low inflation and low unemployment is clearly a trade-off, because they are negatively related to each other.

A final important point is to note that this section derived the Phillips curve from the AS curve, meaning that the two are two sides of the same coin. The reader may equally refer to the AD-AS model as the AD-Phillips curve model.

7.3.1 Adaptive Expectations and the Non-accelerating Inflation Rate of Unemployment (NAIRU)

Perhaps the simplest assumption for π^e_{t-1} is that consumers expect inflation to persist at its current rate. That is, if inflation printed at, say, 4% last year, then consumers expect that it will print at 4% again. Then, $\pi^e_{t-1} = \pi^{\text{act}}_{t-1}$, and so Eq. (7.16) can be re-written as

$$\pi^{\text{act}}_t = \pi^{\text{act}}_{t-1} - \beta(u_t - \bar{u}) + \varepsilon_t. \tag{7.17}$$

It is now clear that if $u = \bar{u}$ then $\mathbb{E}_{t-1}[\pi^{\text{act}}_t] = \pi^{\text{act}}_{t-1}$. That is, inflation is expected to remain steady. For this reason, $\bar{u}$ is called the non-accelerating inflation rate of unemployment or NAIRU.

7.4 The Quantity Theory of Money and Monetarism

Let T denote the number of transactions in the economy in a given year and P denote the average price per transaction. PT is therefore the total amount of money paid for goods and services during the year. For a money supply M, the number of times that this money supply changes hands during the year is therefore,

$$V = \frac{PT}{M}.$$

V is known as the velocity of money. Since T is difficult to measure in practice, economists replace it with income Y. Substituting this into the above equation, and rearranging gives us,

$$MV = PY. \tag{7.18}$$

Equation (7.18) is known as the quantity theory of money, although the reader may note that at this stage, we have an identity, not a theory. However, in the next subsection, we see that this identity is central to the monetarist approach.

7.4.1 Monetarism

Friedman and Schwartz (1963) noted the correlation between M and PY. This drove the monetarist view of inflation. In this view, V is not necessarily constant, but it is close. It varies predictably with the business cycle, and its variation can be adjusted for my policymakers. By pinning down V, Eq. (7.18) becomes a theory of inflation. If V is stable then it can be interpreted as saying that the central bank, which controls the money supply, ultimately controls inflation. If the central bank increases the money supply faster than Y increases, then the price level P rises. Monetarists therefore typically advocate stable or consistent growth in the money supply in line with the natural growth rate in the economy.

Milton Friedman's famous quotation summarises the theory,

> Inflation is always and everywhere a monetary phenomenon in the sense that it is and can be produced only by a more rapid increase in the quantity of money than in output.

However, for all its emphasis on money supply, like the IS-LM model, monetarism does not clearly define what constitutes money. For instance, is money currency, M1, M2, M3, bank reserves, another aggregate, or a combination of these? If government bills and reserves both earn a similar interest rate, then do government bills count as money? Owning bills is not that different to owning bonds since a bill is just a short maturity bond. If government bills are money, then is all government debt money? Unfortunately, $MV = PY$ is silent on this.

The QE programs in the aftermath of the GFC have resulted in large increases in bank reserves, an aggregate that monetarists may have suspected would drive inflation according to $MV = PY$. This has not happened, and instead the post GFC (pre Coronavirus) period has been one of low and stable inflation, with inflation in economic areas that have implemented large QE programs, such as Europe, Japan and the U.S. averaging well below 2% and the growth rate of central bank liabilities.

The FTPL, discussed later in the chapter provides some insight. It argues that it is the total liability of the government (relative to future taxes) that matters for inflation, whether that is cash, government bonds or other liabilities. It therefore notes that exchanging reserves for government bonds

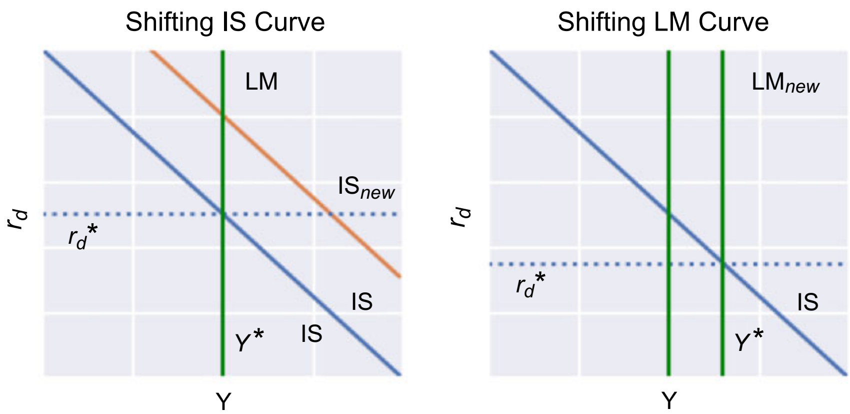


Fig. 7.6 The figure reconciles monetarism with the IS-LM model. The LM curve is vertical. Shifting the IS curve through, for example, a Keynesian fiscal stimulus, has no impact on output. However, if P is sticky then an increase in the money supply shifts the LM curve to the right and this can increase output

does nothing to alter its total liability, and therefore QE is not necessarily inflationary.

I study the FTPL in Sect. 7.5. However, at this stage, the reader may understand a shortcoming of monetarist theory through the following intuition. If the central bank were to double the money supply next Monday, and also announce that it will half it back to its original level the following Monday, presumably the economy would not react, because investors incorporate the expectation that the action will be reversed. The change is temporary. This example shares some similarities with QE. In QE, government bonds that are purchased by the central bank in exchange for newly created reserves will eventually mature. As they do, the reserves are returned to the central bank. The monetarist view, taken in its simplest form, does not incorporate the expectation that eventually the bonds will be paid back. That is, it does not include expectations of future M. Like the AD-AS model, it suffers from not incorporating time consistent behaviours.

7.4.2 Monetarism, IS-LM, and Keynesian Stimulus

Monetarism (steady V) can be thought of as the special case in which the LM curve of Sect. 6.2.1 is vertical. To see this, note that $M/P = \frac{1}{V}Y$. That is, money demand in this model is $L(r_d, Y) = \frac{1}{V}Y$, which is not a function of r_d, and is therefore drawn as a vertical line in the IS-LM diagram. I show this in Fig. 7.6.

The left panel in Fig. 7.6 shows that with a vertical LM curve, a *Keynesian* fiscal stimulus, by which I typically mean an increase in government spending, that shifts the IS curve to the right (as discussed in Sect. 6.2.5) has no impact on output, because the IS and LM curves intersect at the same level of Y. Based on this, monetarists may therefore not support such stabilisation policy.

Monetary stimulus, by which I mean an increase in M, however, may have an impact on output because it increases $P \times Y$. If prices are slow to react and sticky (constant P) then such stimulus can increase Y. In the long run, prices increase, leading M/P to revert back to its previous level, the LM curve shifts back to the left, and Y reverts back to back to its trend level. The right panel in Fig. 7.6 reconciles this with the IS-LM model. An increase in M moves the LM curve to the right. This lowers the interest rate, and increases output.

7.5 Fiscal Theory of the Price Level (FTPL)

The FTPL was pioneered by economist John Cochrane. I refer the reader to Cochrane (2021) for a treatise on this subject. This section provides a short summary of the basic one period, multiperiod and core ideas to allow market practitioners build intuition behind this insightful and intuitive approach to understanding inflation.

The main takeaway is that the FTPL is a backing theory. Just as the stock price of a company is *backed* by the present value of expected future dividends, and their revisions cause the stock price to rise and fall, we will see that the real value of government liabilities is *backed* by the present value of expected real primary surpluses, and their revisions cause the price level in the economy to rise and fall. This should make the FTPL particularly appealing to market practitioners, who are used to calculating prices using present value equations. It means that they can value government debt in a manner that is consistent with how they approach other assets and derivatives.

Further, practitioner adage that governments with excessive debts relative to taxation need to either default, or default via inflation, or/and devalue their FX rate emerges directly and naturally from the FTPL. The reader must keep this in mind when assessing a result that emerges from the FTPL that runs counter to conventional wisdom, that an increase in the nominal interest rate can cause inflation to increase. The argument intuitive in the limit; if nominal interest rates rose to extremely high levels, it is clear that a government would struggle to finance itself, leading to outright default or default via very high inflation. Why then, should it not be true for smaller increases in nominal interest rates to lead to smaller increases in inflation? The FTPL does provide the more conventional result that an increase in the nominal interest rate lowers inflation in the short run, but only by adding features such as long-term debt. However, the unconventional result in the basic form of the FTPL provides practitioners with food for thought. I discuss these ideas over the remainder of this section.

Finally, the FTPL provides at least two further important insights. The first is that the central bank influences inflation expectations. The second is that unexpected inflation, by which I mean realised inflation less expected inflation, comes only from fiscal policy. I show how these ideas emerge from the government's valuation equation.

The reader should note that the FTPL takes expectations into account, making it immune to the Lucas critique, and it incorporates the central pricing equation of Chapter 1, making it consistent with the ideas around the risk premium. Let us begin.

7.5.1 The One-Period Model

Consider a one-period economy consisting of households, goods and services on which households may spend money, cash that pays no interest, and government bonds that pay a positive rate of interest.

Time T is the terminal time for this economy. At time $T-1$, households finish their shopping for that period, make their tax payments and receive transfer payments, and then hold all of their savings in Treasury bonds until time T, because cash pays no interest. At time $T-1$, the government issues a quantity of B_{T-1} bonds to facilitate this saving. Each bond pays 1 USD at time T.

Time T arrives. The government must now pay B_{T-1} USD to households as their bonds mature, collect taxes due and make transfer payments. Let s_T denote the real net primary surplus, by which I mean taxes minus transfers (see Sect. 5.6 for more formal definitions of government accounts). The nominal value of the primary surplus is therefore $s_T P_T$, where P_T is the price level at time T. The net cash that households are left with, which I denote by M_T, is therefore,

$$M_T = B_{T-1} - P_T s_T. \tag{7.19}$$

However, time T is the terminal time in this one-period economy. Households have no need for cash, and so they should purchase goods and services in period T until $M_T = 0$. In doing so, they drive P_T higher. Equation

(7.19) therefore becomes,

$$0 = B_{T-1} - P_T s_T.$$

Rearranging this, we have the simplest form of the FTPL,

$$\frac{B_{T-1}}{P_T} = s_T. \tag{7.20}$$

This equation says that the real value of nominal government debt (the left-hand-side) must equal the real primary surplus (the right-hand side). If surpluses are too small, then the price level rises to make this equation hold. Most importantly, P_T has been determined.

7.5.2 A Multiperiod Model

Next, let us extend the model to multiple periods. There is still no long-term (multiperiod) debt in that bonds mature each period, and the government then issues more. However, we will see that this model provides two key insights. The first is that the central bank controls inflation expectations, and the second is that fiscal policy determines unexpected inflation.

With multiple time periods Eq. (7.19) becomes,

$$Q_t B_t = B_{t-1} - P_t s_t. \tag{7.21}$$

Here, Q_t denotes the nominal price of the bond. This equation is understood as follows. The right-hand side is the same as in Eq. (7.19)—it is the cash that households have left over at time t after collecting B_{t-1} USD from their matured bonds and paying $P_t s_t$ USD in net taxes back to the government. However, in this economy time goes on. Again, the interest rate is positive, and so households do not hold cash into the next period, and instead buy newly issued government debt. Each bond costs Q_t, and B_t of face value is issued hence the left-hand side is $Q_t B_t$.

The final equation that we need is for the nominal bond price,

$$Q_t \equiv e^{-r_t} = \beta_d \mathbb{E}_t^P \left[\frac{P_t}{P_{t+1}} \right], \tag{7.22}$$

where r is the nominal continuously compounded one-period interest rate. Note that β_d is separate from β that was used in the utility functions of Part 1 of this book—I link the two in the (optional) subsection below. Interpreting

β_d as the real discount rate, $\beta_d = e^{-r^{\text{real}}}$, where r^{real} is the real continuously compounded one-period interest rate, Eq. (7.22) is simply the standard Fisher equation that we met in Eq. (3.5). To see this, take logs on both sides to find,

$$r_t = r^{\text{real}} + \pi_t, \tag{7.23}$$

where π_t is expected inflation.[3] Equation (7.22) should therefore be intuitive to market practitioners.

Substituting Eq. (7.22) into Eq. (7.21) we have the following recurrence relation,

$$\frac{B_{t-1}}{P_t} = s_t + \beta_d \mathbb{E}_t^P \left[\frac{B_t}{P_{t+1}} \right]. \tag{7.24}$$

Iterating this equation forward we have our key equation,[4]

$$\frac{B_{t-1}}{P_t} = \mathbb{E}_t^P \sum_{j=0}^{\infty} \beta_d^j s_{t+j}. \tag{7.25}$$

This says that the real value of nominal debt today (the left-hand side), must equal the present value of future real primary surpluses (the right-hand side). The FTPL says that the price level adjusts today so that this equation holds.

Interpretation and Intuition

Equation (7.25) can be interpreted as saying that the government must back its debt with future surpluses. If it does not, then the price level rises today. However, importantly, it is the present value of surpluses that matters. During recessions, governments typically run deficits. Equation (7.25) does not imply that this necessarily causes inflation because the government may run surpluses in the future to offset today's deficit. However, if the markets believe that the government will not or cannot run surpluses in the future, then the price level rises today. This equation shows that the credibility of the government matters for inflation, an observation that matches practitioner intuition.

[3] Note that I have used the approximation that $\ln(1+x) \approx x$ for small x as follows, $\pi_t \equiv \mathbb{E}_t\left[\ln \frac{P_{t+1}}{P_t}\right] = \mathbb{E}_t\left[-\ln \frac{P_t}{P_{t+1}}\right] \approx -\mathbb{E}_t\left[\frac{P_t}{P_{t+1}}\right] + 1 \approx -\ln \mathbb{E}_t\left[\frac{P_t}{P_{t+1}}\right]$.

[4] I assume that the transversality condition, $\lim_{T\to\infty} \mathbb{E}_t\left[\beta_d^T \frac{B_{T-1}}{P_T}\right] = 0$ holds.

Further, the reader may be familiar with the present value formula for valuing stocks,

$$nS_t = \mathbb{E}_t^P \sum_{j=0}^{\infty} \beta_s^j d_{t+j}. \tag{7.26}$$

The left-hand side is the market value of the firm, given by the number of shares outstanding, n, times the price in USD (or other currency) per share, S_t. The right-hand side is expected future dividends, discounted at an appropriate risk-adjusted rate, which I denote by β_s. S_t varies each day as the market re-assesses the discount rate and expected dividends. The dividends *back* the stock price.

Equation (7.25) is analogous. It says that the number of bonds outstanding, B_t, multiplied by their price in terms of goods, $1/P_t$,[5] equals the present value of real goods being collected by the government in the future in net taxes. That is, just as market practitioners calculate present-values for stocks, they should do the same for a government's future surpluses to calculate the price level. Cochrane writes that "nominal government debt is *stock in the government*."

Relationship to Monetarism

In the FTPL, inflation results from one source—more money in the economy than is soaked up by net nominal tax collection. This is different to monetarism, where inflation results from more money than is needed for transactions. Further, the FTPL makes clear that only money issued by the government, and government bonds that promise such money in the future drive inflation. In monetarism, all money counts as money and causes inflation. For example, monetarism must argue that privately issued cryptocurrencies, if adopted for transactions, create inflation.

Relationship to AD-AS

In the AD-AS model, a right shift in the AD curve caused, perhaps, by households consuming more, creates a higher price level. The FTPL views this as a wealth effect on government bonds in the sense that households see money and government bonds that are not sufficiently backed by future surpluses as worthless relative to goods, and thereby they spend more to drive up prices.

[5] P_t is the price of goods in terms of USD, and so $1/P_t$ is the price of USD in terms of goods.

Deriving Equation (7.22) from the Central Pricing Equation (Optional) We can derive Eq. (7.22) via the central pricing equation of Chapter 1. It is Eq. (1.31), with the Q_t^{real} written as Q_t/P_t, and the real payoff written as $1/P_{t+1}$. We then have,

$$Q_t = \beta \mathbb{E}_t^P \left[\frac{u'(c_{t+1})}{u'(c_t)} \frac{P_t}{P_{t+1}} \right] = \beta_d \mathbb{E}_t^P \left[\frac{P_t}{P_{t+1}} \right].$$

Therefore,

$$\beta_d = \beta \frac{\mathbb{E}_t^P \left[\frac{u'(c_{t+1})}{u'(c_t)} \frac{P_t}{P_{t+1}} \right]}{\mathbb{E}_t^P \left[\frac{P_t}{P_{t+1}} \right]}.$$

In the special case of risk neutral investors so that $u'(c_t) = 1$, then $\beta_d = \beta$. However, when investors are not risk neutral then the right-hand side of the above equation can vary. Perhaps $\beta_{d,t}$ would have been more appropriate notation for the left-hand side. Cochrane (2021) provides a full treatment of time varying risk premiums in the FTPL.

7.5.3 Monetary Policy and Expected Inflation

The FTPL argues that monetary policy sets inflation expectations. To understand this, let us derive an equation for expected inflation. First, note that Eq. (7.25) holds at every point in time. I rewrite it at time $t + 1$ as,

$$\frac{B_t}{P_{t+1}} = \mathbb{E}_{t+1}^P \sum_{j=0}^{\infty} \beta_d^j s_{t+1+j}. \tag{7.27}$$

Next, take the expectation on both sides at time t,[6]

$$\mathbb{E}_t \left[\frac{B_t}{P_{t+1}} \right] = \mathbb{E}_t^P \sum_{j=0}^{\infty} \beta_d^j s_{t+1+j}. \tag{7.28}$$

In words, this equation says that the expectation of the real value of nominal debt one period in the future is equal to the present value of real surpluses

[6] The right-hand side follows from the law of iterated expectations, which says that $\mathbb{E}_t[X] = \mathbb{E}_t[\mathbb{E}_{t+1}[X]]$ for a random variable X.

one period into the future. Finally, we arrive at

$$\underbrace{\frac{B_t}{P_t}e^{-r_t}}_{\text{Term 1.}} = \underbrace{\frac{B_t}{P_t}\beta_d \mathbb{E}_t^P\left[\frac{P_t}{P_{t+1}}\right]}_{\text{Term 2.}} = \underbrace{\mathbb{E}_t^P \sum_{j=0}^{\infty} \beta_d^{j+1} s_{t+1+j}}_{\text{Term 3.}} . \tag{7.29}$$

Term 2 is obtained by substituting in Eq. (7.22), and term 3 is obtained by substituting in Eq. (7.28).

Let us analyse this equation. P_t is the price level, and this is already set at time t. Therefore, the remaining variables that are able to change are B_t, r_t, and $\mathbb{E}_t[1/P_{t+1}]$.

The central bank sets monetary policy, by which I mean that it chooses the nominal interest rate, r_t. However, it does not influence surpluses because this is the job of fiscal policy, meaning that term 3 is constant. Therefore, Eq. (7.29) says that the quantity of debt that the government may sell at time t for maturity at time $t + 1$, B_t, is determined by the nominal interest rate that the central bank sets. If r_t is set at a high level, then B_t must rise to ensure that term 1 remains equal to term 3. That is, if the central bank sets a high-interest rate, then it can sell a larger quantity of debt.

However, term 2 provides another implication. If B_t is large, then $\mathbb{E}_t\left[\frac{P_t}{P_{t+1}}\right]$ must fall to offset this, to ensure that term 2 remains equal to term 3. The same surpluses are backing a larger quantity of debt, and so expected inflation rises. The implication is that a higher interest rate raises expected inflation!

Practitioners may be unconvinced, with conventional wisdom being that a rise in interest rates decreases inflation. However, the logic here is that a higher nominal interest rate lowers Q_t in Eq. (7.21). The government therefore rolls over its debt at a worsening price. Further, Eq. (7.25) tells us that a worsening net worth of the government raises the price level.[7]

Here, I remind the reader of the relatively uncontroversial argument that a government that cannot finance its liabilities must default explicitly or via inflation. Being unable to roll over debt is the same as saying that the interest rate is too high. In the limit of very high-interest rates, therefore, practitioners are comfortable with the implication being very high inflation. It is therefore

[7] Note that here *net worth* refers to a hypothetical government balance sheet that includes all expected future surpluses and liabilities. Such items do not appear in the IMF summary of government balance sheets discussed in Chapter 5.

not unreasonable to suggest that smaller increases in interest rates may lead to smaller increases, but increases nonetheless, in inflation.

Long term debt

The model has so far assumed that the government is rolling over one-period debt. Adding long-term debt can provide the opposite and more common result that raising interest rates can lower inflation today. To see this, generalise Eq. (7.25) to

$$\frac{\sum_{j=0}^{\infty} Q_t^{t+j} B_{t-1}^{t+j}}{P_t} = \mathbb{E}_t^P \sum_{j=0}^{\infty} \beta_d^j s_{t+j}. \tag{7.30}$$

Here, B_{t-1}^{t+j} denotes number of bonds (each paying 1 USD) maturing at time $t+j$ and Q_t^{t+j} is the nominal price per bond. Q_t^{t+j} is calculated by generalising Eq. (7.22) to

$$Q_t^{t+j} = e^{-r_{t,t+j}j} = \mathbb{E}_t^P \left[\beta_d^j \frac{P_t}{P_{t+j}} \right], \tag{7.31}$$

where $r_{t,t+j}$ is the continuously compounded nominal market interest rate from time t to $t+j$. Equation (7.30) now says that the real *market* value of nominal debt equals the present value of real surpluses.

Next, suppose that the central bank unexpectedly raises the interest rate and provides forward guidance that it will stay high forever. That is, $r_{t,t+j}$ rises for every j. Assuming no change in expected surpluses, the right-hand side remains fixed. However, we can see from Eq. (7.31) that longer dated debt falls in value more than shorter dated debt. Practitioners know this as the duration effect.

If the government has funded itself with mainly longer dated debt, meaning that B_{t-1}^{t+j} is larger for larger values of j, then the numerator, $\sum_{j=0}^{\infty} Q_t^{t+j} B_{t-1}^{t+j}$, falls. Therefore, P_t must fall today. That is, a *hawkish* central bank announcement leads to a lower price level today.

The intuition is that the government's net worth benefits from having issued long-term bonds, because it did so when the nominal interest rate was lower, and therefore received a higher price for them. In practitioner parlance, the government has a *paid position* in interest rates and this position improves its market value with unexpectedly higher interest rates. This caused the price level to fall today, but assuming that the interest rate stays high, it will eventually cause inflation to rise in the future because the government will eventually need to refinance at a higher interest rate. We see this in Eq. (7.31) where lower nominal bond prices Q_t^{t+j} imply higher future inflation because $\mathbb{E}_t^P[P_t/P_{t+1}]$ must be lower.

Risk Premiums

Section 1.6.2 discussed how the real interest typically falls during times of economic turmoil, as investors revise down expectations of growth, and carry greater precautionary savings. Recall that $\beta_{d,t} = e^{-r_t^{\text{real}}}$, where I have added the t subscript to allow the real interest rate to vary with time. If the real interest rate is lower, then the present value of surpluses is higher because $\beta_{d,t}$ is higher. That is, the right-hand side of Eq. (7.30) is higher. The left-hand-side must increase to match. It can do this either through a deflation (fall in P_t), or a rise in the market price of bonds, or a combination of these impacts.

7.5.4 Fiscal Policy and Unexpected Inflation

Next, let us derive an equation for unexpected inflation. First, multiply and divide Eq. (7.27) by P_t,

$$\frac{B_t}{P_t}\frac{P_t}{P_{t+1}} = \mathbb{E}_{t+1}\sum_{j=0}^{\infty}\beta_d^j s_{t+1+j}. \tag{7.32}$$

Therefore,

$$\begin{aligned}\frac{B_t}{P_t}\underbrace{\left(\frac{P_t}{P_{t+1}} - \mathbb{E}_t\left[\frac{P_t}{P_{t+1}}\right]\right)}_{\text{Term 1.}} &= \mathbb{E}_{t+1}\sum_{j=0}^{\infty}\beta_d^{j+1}s_{t+1+j} - \mathbb{E}_t\sum_{j=0}^{\infty}\beta_d^{j+1}s_{t+1+j}.\\ &= \beta s_{t+1} + \mathbb{E}_{t+1}\sum_{j=1}^{\infty}\beta_d^{j+1}s_{t+1+j}\\ &\quad - \mathbb{E}_t\sum_{j=0}^{\infty}\beta_d^{j+1}s_{t+1+j}. \end{aligned}\tag{7.33}$$

Term 1 is the difference between the (inverse) of the inflation that is realised between time t and time $t+1$, and what was expected at time t. The right-hand side is the difference between the present value of expected surpluses calculated at time $t+1$ and that at time t. At time $t+1$ we have observed s_{t+1} and therefore this term can come out of the expectation.

The important take away is that if there is an improvement in the present value of expected surpluses, then inflation undershoots its ex-ante expectation, and vice versa if there is a worsening. Unexpected inflation is therefore controlled by fiscal policy and not by the central bank. The government does not have to run a surplus to lower unexpected inflation. For example, during the Coronavirus pandemic of 2020, s_{t+1} was negative as governments enacted support packages for households negatively impacted by the closure of parts of the economy. However, if $\mathbb{E}_{t+1} \sum_{j=1}^{\infty} \beta_d^{j+1} s_{t+1+j}$ rises by more than βs_{t+1} is negative, then the result is lower inflation. It is likely that only a credible government will be able to achieve this and this example re-emphasises the importance of fiscal credibility for controlling inflation.

7.5.5 A Note on QE

Recall that QE involves the central bank purchasing government bonds from the private market in exchange for reserves. The FTPL argues that this is not inflationary. By exchanging one type of government liability (government bonds) for another (reserves held at the central bank), the total liabilities of the broad government, $\sum_{j=0}^{\infty} Q_t^{t+j} B_{t-1}^{t+j}$ in Eq. (7.30), is unchanged.[8] The operation does not impact surpluses and so the right-hand side of Eq. (7.30) is also unchanged. Therefore, this action should not impact the price level.

Although the mechanism through which QE impacts the economy is poorly understood by central bankers (see, for example, "BoE criticised by internal watchdog over easing programme" (*Financial Times*, 2021) for the BOE's own admission of such), it remains an active policy tool. However, note that if we do take it as given that the impact on the economy is positive then this should increase the present value of surpluses. This effect is deflationary, not inflationary, in the FTPL.

Separately, recall from Sect. 5.6.3 that QE shortens the maturity structure of government debt held by the private sector. As discussed in Sect. 7.5.3, this may change the reaction of the price level to movements in the interest rate. With shorter term debt outstanding, inflation is *more* positively correlated with the nominal interest rate.

[8] If the transaction were carried out at the mid-market rate, then there is no change in this quantity. However, the trade may carry a price impact, which may change this quantity, causing some impact on the price level.

7.5.6 A Note on Helicopter Money

Helicopter money refers to newly printed money that is distributed to households. The main difference between helicopter money and QE is that the latter involves the central bank adding both an asset (government bonds) and a liability (reserves) to its balance sheet, whereas with the former, the government adds a liability only in that there is more money outstanding, but there is no explicit offsetting asset.

The commonly made inflationary argument for helicopter money is that households spend it, which increases demand, pushes consumer prices higher and stimulates the economy. However, in the FTPL, whether there is inflation or not depends on whether expected surpluses rise to offset the stimulus. According to the theory, if the government intends to raise taxes in the future then there is no need for the price level to adjust today. To reconcile this with household behaviour, consider the following. If consumers expect that they must repay the helicopter money in future taxes, then they should save the helicopter money today and there therefore no stimulatory effect. We described such a government as Ricardian in Sect. 6.2.6. In this case, although it is not explicitly recorded, there is indeed an additional asset on the government's balance sheet: future tax collections.

Helicopter money is therefore inflationary only if the present value of real expected surpluses does not rise in value to sufficiently offset the real value of the newly created money. That is, the fiscal authority is non-Ricardian. To illustrate the idea, consider hyperinflationary periods, such as Germany in the 1920s, Hungary in the 1940s and Zimbabwe in the 2000s, among others. It is clear that the fiscal authority in these circumstances was non-Ricardian.

7.6 New-Keynesian Phillips Curve (NKPC) Models and Taylor Rules

NKPC models share characteristics with *old* Keynesian IS-LM models, but with the important feature that they incorporate expectations and thereby hope to survive the Lucas critique. Such models form a large part of academic work on monetary theory and central bank models. Here, I provide a short tour of the important ideas and refer the reader to Woodford (2003) for a treatise on this topic.

7.6.1 The New-Keynesian IS Equation

The (simplest form of) new-Keynesian IS equation is

$$y_t - \bar{y} = \mathbb{E}_t^P[y_{t+1} - \bar{y}] - \lambda(r_t^{\text{real}} - r^*). \tag{7.34}$$

Here, y_t is the log of output at time t, $y_t = \ln Y_t$, $\bar{y}$ is the log of the natural level of output $\bar{y} = \ln \bar{Y}$, and r^* is the natural real interest rate. I return to r^* later in this section. Note that texts often denote $\bar{y}$ and r^* as $\bar{y}_t$ and r_t^* to capture the slow variation over time of these quantities.

Inspecting Eq. (7.34) the reader may note the similarity with the IS equation of the IS-LM model in that a rise in the real interest rate (above r^*) causes output to fall (below its natural level).

Equation (7.34) is not a new introduction. It is simply a re-expression of the price of the real bond which we met in Eq. (1.32). To see this, write the real bond price as Q_t/P_t and the real payoff as $\frac{1}{P_{t+1}}$. Equation (1.32) then becomes

$$\frac{Q_t}{P_t} = \mathbb{E}_t^P\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\frac{1}{P_{t+1}}\right], \tag{7.35}$$

where $Q = e^{-r}$. Next, take logs on both sides and again approximate $\ln \mathbb{E}[X] = \mathbb{E}[\ln X]$ to find,

$$-r_t - \ln P_t = \mathbb{E}\big[\ln\beta - \gamma(\ln c_{t+1} - \ln c_t) - \ln P_{t+1}\big]. \tag{7.36}$$

Next, substituting in $\pi_{t+1} = \ln P_{t+1} - \ln P_t$ and rearranging we have,

$$\mathbb{E}_t^P\big[\ln c_{t+1} - \ln c_t\big] = \frac{\ln\beta}{\gamma} + \frac{1}{\gamma}(r_t - \mathbb{E}_t^P[\pi_{t+1}]). \tag{7.37}$$

Finally, assuming that consumption c_t equals output Y_t we may write,

$$\mathbb{E}_t^P[(y_{t+1} - \bar{y}) - (y_t - \bar{y})] = \lambda(r_t^{\text{real}} - r^*), \tag{7.38}$$

where $r^* = -\ln\beta$ and $\lambda = \frac{1}{\gamma}$. Rearranging this gives us Eq. (7.34).

In Sect. 1.6.2 we interpreted Eq. (7.35) as saying that investors form expectations of the distribution of future consumption, and then choose how much to consume or save in bonds today. The free market demand for saving determined the real interest rate. Here, the interpretation is reversed. We will see in Sect. 7.6.3 that the real interest rate will be controlled by the central bank via a policy rule, and this in turn will determine current consumption and therefore output via Eq. (7.34).

7.6.2 The New-Keynesian Phillips Curve

The (simplest form of the) new-Keynesian Phillips curve is

$$\pi_t = \beta \mathbb{E}_t^P[\pi_{t+1}] + \kappa(y_t - \bar{y}). \tag{7.39}$$

This equation says that the inflation rate today depends on expected future inflation and the output gap. κ captures price stickiness. A lower value of κ implies that prices are less able to change and are therefore more sticky relative to the economic cycle.

The intuition underlying this equation is similar to that in the AD-AS model; inflation is pro-cyclical in that its high inflation corresponds to a positive output gap, and vice-versa. However, if households expect higher inflation tomorrow, then inflation will increase today for any given level of the output gap.

The Phillips curve is most often discussed in industry and media reports as a relationship between inflation and unemployment. This is simply a substitution of Okun's law, $y_t - \bar{y} = -2 \times (u_t - \bar{u})$, as we did in Eq. (7.15).

The Phillips curve remains central in debates on inflation. However, in recent times this debate has focussed on whether the Phillips curve is *alive* or *dead*. By this, commentators refer to the fact that the relationship between unemployment and inflation has been steadily weakening over time. This is often described as a *flattening* in the Phillips curve and it is equivalent to a decrease in κ.

The reason for the flattening Phillips curve is not clear. One commonly cited reason is lower employee wage bargaining power as a result of the decline of unionisation and/or increased globalisation in the labour market. However, there may also be measurement errors in calculating the output gap. Further, forces such as technological progress and its associated productivity improvements tend to decrease prices and are not captured in Eq. (7.39).

7.6.3 The Taylor Rule

The final component is a rule for setting monetary policy. The Taylor rule, named after economist John Taylor, prescribes that the central bank implement a rules-based policy rate given by

$$r_t = \pi_t + r^* + a_\pi(\pi_t - \pi_t^*) + a_y(y_t - \bar{y}). \tag{7.40}$$

Here, a_π and a_y are both positive constants, and π_t^* denotes the central bank's inflation target, typically around 2% per annum. By specifying $a_\pi > 0$, the rule prescribes that an increase in π_t of 1% must be met with a central bank response of more than a 1% increase in the policy rate. Further, a rise in output beyond natural level be must be met with a higher policy interest rate. In his seminal work, Taylor (1993) proposed setting a_π and $a_y = 0.5$.

Rearranging Eq. (7.40) we have,

$$r_t^{\text{real}} - r^* = a_\pi(\pi_t - \pi_t^*) + a_y(y_t - \bar{y}). \tag{7.41}$$

With the policy rule in place we now have a full description of the approach towards controlling inflation in the NKPC model. If $\pi_t > \pi_t^*$ then the central bank raises the policy rate, and the real interest rate rises above the natural rate, $r_t^{\text{real}} - r^* > 0$. This causes downward pressure on the output gap, $y_t - \bar{y}$, via the IS curve, Eq. (7.34), and this downward pressure then causes inflation to fall via the NKPC in Eq. (7.39).

The Natural Rate of Interest, r^*

Laubach and Williams (2003) define the natural rate of interest as the real interest rate consistent with output equaling its natural rate and stable inflation. In Eq. (7.38) we set $r^* = -\ln\beta$. The reader may recall from Eq. (1.34) that $-\ln\beta$ was the *impatience* term in the expression for the real interest rate. This link allows the reader to assign a meaning to the concept of r^* as a component of the consumer's utility function.

7.7 Chapter Summary

- Measuring inflation is challenging. The CPI is a Laspeyres index and thereby tends to overstate inflation. The GDP deflator is a Paasche index, and may understate inflation. The PCE, which is also the Federal Reserve's preferred model, is a Paasche index, and offers a middle ground.
- The AD-AS model has been the workhorse of policy maker and practitioner thinking about inflation for an extended period. The AD curve is

built on the IS-LM/Keynesian model. The AS curve and the Phillips curve are two sides of the same coin.

- There are two *types* of inflation in the AD-AS model; demand-pull inflation, which refers to the AD curve moving to the right, and cost-push inflation, which refers to the AS curve moving to the left.
- Monetarism is built on the quantity theory of money identity. It argues that the velocity of money, V, is stable and therefore the quantity of money must grow at the same rate as output, otherwise the price level changes. Too much money relative to output growth leads to inflation. However, monetarism does not fully define what it means by money.
- The FTPL is a backing theory. It argues that government liabilities need to be backed by expected future surpluses. If surpluses are too small relative to liabilities, then the price level rises.
- The FTPL argues that the central bank controls expected inflation, but unexpected inflation is determined by fiscal policy.
- The FTPL implies that higher interest rates worsen the financial position of the government, and this leads to inflation, all else remaining equal. If there is long-term debt outstanding, then there may be an initial deflation.
- The NKPC approach understands inflation as follows. The central bank credibly commits to a Taylor rule and thereby anchors inflation expectations. If inflation rises above its target, it raises the real interest rate. This lowers output via the IS equation, which in turn lowers inflation today via the NKPC.

References

Cochrane, J. H. (2021). *Fiscal theory of the price level—Web draft*, Print Version Forthcoming.

Financial Times. (2021). Boe criticised by internal watchdog over easing programme. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm

Friedman, M., & Schwartz, A. J. (1963). *A monetary history of the United States, 1867–1960*. Princeton University Press.

Mankiw, N. (2007). *Macroeconomics* (6th ed.). Worth Publishers.

Taylor, J. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*, 195–214.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

Appendix A: Exchange Rate Concepts

Exchange Rate Quotation

The quotation convention in FX markets flows from left to right. For example, EUR-USD trading at 1.20 means that 1 Euro (the currency on the left) is exchanged for 1.20 U.S. Dollars (the currency on the right). Similarly, if USD-JPY is trading at 110 then 1 U.S. Dollar is exchanged for 110 Japanese Yen.

If the exchange rate rises then the currency on the left-hand side is strengthening. For example, if EUR-USD rises from 1.20 to 1.25 then the Euro has strengthened because it now buys 1.25 U.S. Dollars. However, if USD-JPY rises from 110 to 115 then the U.S. Dollar is strengthening because it now buys 115 Japanese Yen.

FX quotation conventions can be confusing, but they are no different in principle to stocks quotation. For example, since the S&P500 index is denominated in U.S. Dollars, if its price is 2000 U.S. Dollars, one could write SPX-USD = 2000, meaning one unit of the S&P500 costs 2000 U.S. Dollars. For EUR-USD one can think of EUR as the stock and USD as the currency that the buyer pays in. For USD-JPY one can think of the USD as the stock and JPY as the currency that the buyer pays in.

FX rates are often quoted in *pips*. A pip or *price interest point* is the smallest price move that a given exchange rate makes based on market convention. So, for example, if EUR-USD moves from 1.20 to 1.2010, we say that it has

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9

moved by 10 pips because market convention is to quote EUR-USD to four decimal places. However, if USD-JPY moves from 110.00 to 110.10, then this is still 10 pips because market convention is to quote USD-JPY to two decimal places.

Appendix B: Probability

Probability Mass Functions (PMFs), Probability Density Functions (PDFs) and Calculating Expectations

Discrete Random Variables and Probability Mass Functions

Suppose that the EUR-USD spot rate at a future time T, S_T, is a random variable that can take one of 3 discrete values, 1.15, 1.20 and 1.25. The probabilities with which these values occur are

$$\begin{aligned}
\mathrm{P}(S_T = 1.15) &= 0.2 \\
\mathrm{P}(S_T = 1.20) &= 0.5 \\
\mathrm{P}(S_T = 1.25) &= 0.3.
\end{aligned}$$

As required, these probabilities sum to 1.

The PMF belonging to S_T is shown in Fig. B.1. It is a plot of the probability with which EUR-USD takes each of its possible values at T. I have shown 3 states here, but we can extend this idea to many discrete states.

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9

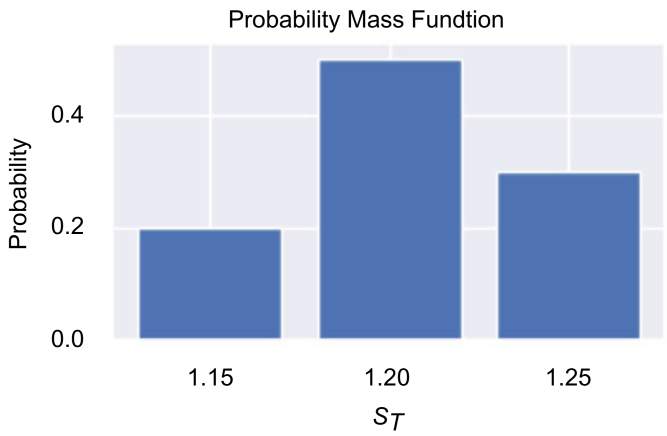


Fig. B.1 The figure shows the PMF of S_T. In this case, S_T can take one of three discrete values. The PMF gives us the probability that S_T takes each of these values

The expected value of S_T, denoted $\mathbb{E}[S_T]$ is calculated as follows.

$$\begin{aligned}\mathbb{E}[S_T] =& \sum_{x \in X} x \Pr(S_T = x)\\ =& \Pr(S_T = 1.37) \times 1.37 + \Pr(S_T = 1.39) \times 1.39\\ &+ \Pr(S_T = 1.41) \times 1.41\\ =& 0.2 \times 1.37 + 0.5 \times 1.39 + 0.3 \times 1.41\\ =& 1.3920,\end{aligned} \tag{B.1}$$

where $X = \{1.37, 1.39, 1.41\}$. In words, the expectation of S_T is the weighted average value of S_T, with the weights given by the PMF.

If we wished to extend this idea to infinitely many, or a continuum of, states then we must turn to PDFs. This is the topic of the next subsection.

Continuous Random Variables and PDFs

Next, suppose that S_T can take a continuum of values. The probability that S_T is between value a and b at time T is given by

$$\Pr(a < S_T \leq b) = \int_a^b f_{S_T}(x)\mathrm{d}x, \tag{B.2}$$

where $f_{S_T}(x)$ is the probability density function belonging to S_T. In words, there is a function $f_{S_T}(x)$, and the area underneath this function between a and b provides the probability that S_T will lie between a and b. Figure B.2 and its caption explain this idea.

The cumulative density function, or CDF, is denoted by $F_{S_T}(b)$ and it is the probability that $S_T \leq b$. That is,

$$F_{S_T}(b) = \Pr(S_T \leq b) = \int_{-\infty}^{b} f_{S_T}(x)\mathrm{d}x.$$

In the case of the normal distribution, which is commonly used in FX markets as a starting approximation, the standard deviation is given by the points a and b such that 68% of the probability is contained. Points a and b are also each equidistant from the center of the distribution. This is shown in Fig. B.3.

Analogous to the PMF, the expected value of S_T is calculated as follows,

$$\mathbb{E}[S_T] = \int_{-\infty}^{\infty} x f_{S_T}(x)\mathrm{d}x.$$

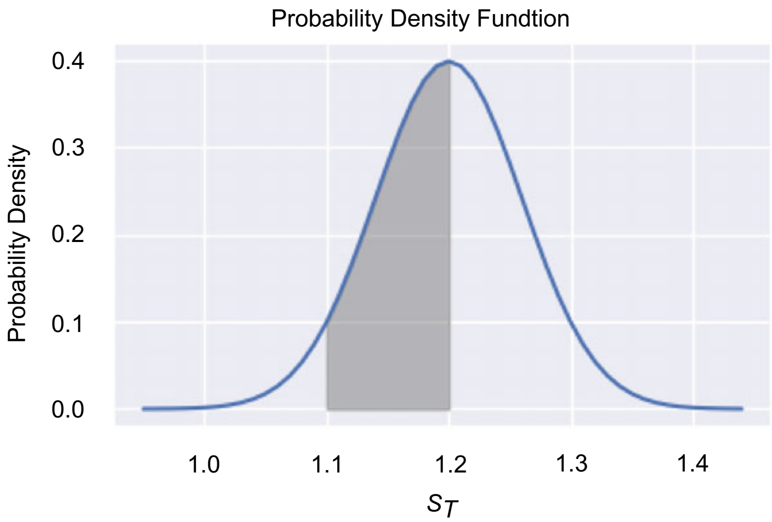


Fig. B.2 The figure shows the probability density function of S_T, $f_{S_T}(x)$. Suppose we wish to calculate the probability that S_T is between $a = 1.1$ and $b = 1.2$ at time *T*. This is found by calculating the shaded area under the PDF, $\int_a^b f_{S_T}(x)\mathrm{d}x$

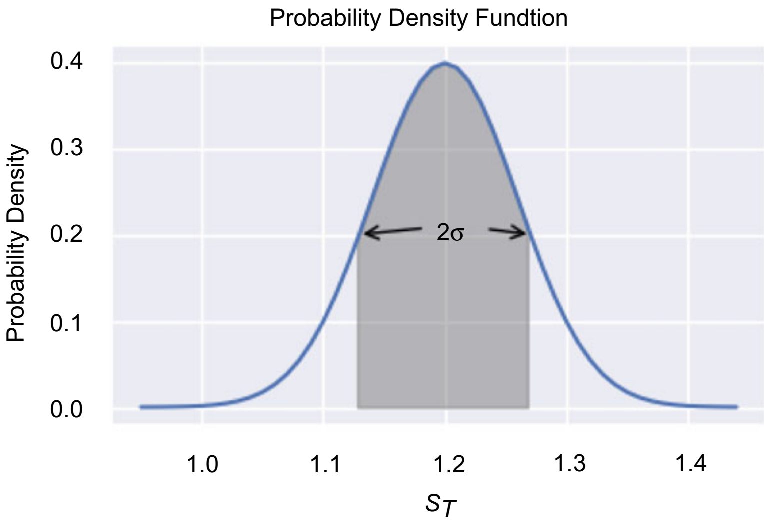


Fig. B.3 The figure shows a normal distribution with a 1 year standard deviation of 6%. This means that the shaded area contains 68% of the probability density. The vertical bars are at $1.20 - 6\% = 1.1280$ and at $1.20 + 6\% = 1.2720$, respectively

Appendix C: Calculus

Partial Derivatives

This section provides a short refresher of the concept of a partial derivative, illustrated with a function of three variables. Denote the function that we are interested in by $V(S, \sigma, \tau)$. For example, this function could be the option valuation function, where S is the spot FX rate, σ is the option implied volatility and τ is the time to maturity. The user plugs the values of variables S, σ and τ into the function V to return the option price. The partial derivative allows us to understand how V changes if any one of S, σ or τ changes, while the other two variables remain constant.

Suppose that initially $S = S_1$, $\sigma = \sigma_1$ and $\tau = \tau_1$. The change in the option value ΔV due to S moving from S_1 to $S_1 + \delta$ is

$$\Delta V = V(S_1 + \delta, \sigma_1, \tau_1) - V(S_1, \sigma_t, \tau_1).$$

Therefore, the change in the option price per unit move in S is

$$\frac{\Delta V}{\delta} = \frac{V(S_1 + \delta, \sigma_1, \tau_1) - V(S_1, \sigma_1, \tau_1)}{\delta}. \tag{C.1}$$

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9

Taking the limit as $\delta \to 0$ we have the *partial* derivative of the option price with respect to S, evaluated when $S = S_1$.

$$\left.\frac{\partial V(S,\sigma,\tau)}{\partial S}\right|_{S=S_t} \equiv \lim_{\delta \to 0} \frac{V(S_1+\delta,\sigma_1,\tau) - V(S_1,\sigma_1,\tau_1)}{\delta}.$$

If S moves from S_1 to a new level S_2 while σ_1 and τ_1 remain unchanged then the new value of the option is

$$V(S_2,\sigma_t,\tau) = V(S_1,\sigma_1,\tau_1) + \left.\frac{\partial V(S,\sigma,\tau)}{\partial S}\right|_{S=S_1} \times (S_2 - S_1) + \text{ HOT}, \tag{C.2}$$

where HOT refers to *higher order terms*. This approximation is generally more accurate as $S_2 - S_1$ gets smaller.

There are at least three important points to remember. First, the partial derivative provides the sensitivity of the function to one of the variables changing, while holding the others fixed. Above, we allowed S to change, but we could equally have allowed σ or τ to change. Second, the partial derivative is accurate to first order, or infinitesimally small moves in S only. For larger moves, one must treat the above calculation as an approximation only, or include higher order terms. Third, the partial derivative is itself (in general) a function of the variables (S and σ in the example above). It is therefore important to note where the partial derivative is evaluated.

Finally, Eq. C.2 is often written in the form

$$\mathrm{d}V = \frac{\partial V(S,\sigma,\tau)}{\partial S}\mathrm{d}S. \tag{C.3}$$

Here, $\mathrm{d}S$ and $\mathrm{d}V$ denote infinitesimal changes in S and V, respectively. The partial derivative is therefore simply the ratio of the change in V when S moves by an infinitesmal amount.

Total Derivatives

Section C.1 studied the situation in which S moved. Let us now consider a circumstance in which S, σ and τ all change. How can we calculate the change in $V(S,\sigma,\tau)$? The answer is to calculate the total derivative as

$$dV = \frac{\partial V(S, \sigma, \tau)}{\partial S} dS + \frac{\partial V(S, \sigma, \tau)}{\partial \sigma} d\sigma + \frac{\partial V(S, \sigma, \tau)}{\partial \tau} d\tau. \tag{C.4}$$

Here, we have simply chained together the partial derivatives. The equation says that if S, σ and τ each change by infinitesimal amounts dS, $d\sigma$ and $d\tau$, then the change in V, dV is given by these changes multiplied by the respective partial derivatives.

Glossary of Acronyms

AD Aggregate Demand
AS Aggregate Supply
ATM At-the-money
ATMF At-the-money-forward
AUD Australian Dollar
BLS Bureau for Labour Statistics
BOE Bank of England
BOJ Bank of Japan
BOP Balance of Payments
BSM Black-Scholes-Merton
CAPM Capital Asset Pricing Model
CDF Cumulative Distribution Function
CDS Credit Default Swap
CCB Cross Currency Basis
CHF Swiss Franc
CIP Covered Interest Parity
CPI Consumer Price Index
CRRA Constant Relative Risk Aversion
DM Developed Market
DNS Delta Neutral Straddle
ECB European Central Bank
EDP Economic Decimation Party
EGP Economic Growth Party
EM Emerging Market
EMH Efficient Markets Hypothesis

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9

EUR Euro
FOMC Federal Open Markets Committee
FRB Federal Reserve Board
FTPL Fiscal Theory of the Price Level
FX Foreign Exchange
GBP British Pound
GDP Gross Domestic Product
GFC Global Financial Crisis
GNDI Gross National Disposable Income
GSG Global Savings Glut
HOT Higher Order Terms
IIP International Investment Position
IMF International Monetary Fund
JGB Japanese Government Bond
JPY Japanese Yen
MXN Mexican Peso
NAIRU Non-Accelerating Inflation Rate of Unemployment
NIIP Net International Investment Position
NIPA National Income and Product Accounts
NKPC New Keynsian Philips Curve
NOB Net Operating Balance
NOK Norwegian Krone
OCA Optimal Currency Area
PBOC People's Bank of China
PCE Personal Consumption and Expenditures
PDF Probability Density Function
PIB Primary Income Balance
pip Price Interest Point
PPP Purchasing Power Parity
QE Quantitative Easing
QOQ Quarter-Over-Quarter
REER Real Effective Exchange Rate
RER Real Exchange Rate
SIB Secondary Income Balance
SDR Special Drawing Rights
SGO Statement of Government Operations
SNB Swiss National Bank
S&P Standard and Poor's
TRY Turkish Lira
TIPS Treasury Inflation Protected Securities
UIP Uncovered Interest Parity
U.K. United Kingdom
U.S. United States
USD U.S. Dollar

VaR Value-At-Risk
VRP Volatility Risk Premium
YCC Yield Curve Control
YOY Year-Over-Year

References

Balassa, B. (1964). The purchasing power parity doctrine: A reappraisal. *Journal of Political Economy, 72*, 584.

Bernanke, B. S. (2005). *The global saving glut and the US current account deficit*. Published on website of the FRB.

Bernanke, B. S., Bertaut, C., DeMarco, L. P., & Kamin, S. (2011). International Capital Flows and the Returns to Safe Assets in the United States, 2003–2007. *FRB International Finance Discussion, 15*, 13–26.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637.

Borio, C., Iqbal, M., McCauley, R., McGuire, P., & Sushko, V. (2018). *The failure of covered interest parity: FX hedging demand and costly balance sheets* (BIS Working Papers, No. 590).

Brandt, M., Cochrane, J. H., & Santa-Clara, P. (2006). International risk sharing is better than you think, or exchange rates are too smooth. *Journal of Monetary Economics, 53*(4), 671–698.

Cochrane, J. H. (2005). *Asset Pricing* (Rev. ed.). Princeton University Press.

Cochrane, J. H. (2011). Discount rates. *Journal of Finance, 66*, 1047–1108.

Cochrane, J. H. (2021). *Fiscal theory of the price level—Web draft*, Print Version Forthcoming.

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics, 7*, 229–264.

Eichengreen, B. (2011) *Exorbitant privilege: The rise and fall of the dollar*. Oxford University Press.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417.

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9

Financial Times. (2021). Boe criticised by internal watchdog over easing programme. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm

Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.

Friedman, M., & Schwartz, A. J. (1963). *A monetary history of the United States, 1867–1960*. Princeton University Press.

Gopinath, G., & Stein, J. C. (2018). *Banking, trade and the making of a dominant currency* (NBER Working Paper, No. 24485).

Hicks, J. R. (1937). Mr. Keynes and the classics: A suggested interpretation. *Econometrica, 5*, 147–159.

Iqbal, A. (2018). *Volatility*. Wiley Finance Series: Practical Options Theory.

Keynes, J. M. (1936). *The general theory of employment: Interest and money*. Macmillan.

Krugman, P. (2007). Will there be a dollar crisis? *Economic Policy, 22*(51), 435–467.

Lane, P. R. (2012). The European Sovereign debt crisis. *Journal of Economic Perspectives, 26*(3), 49–68.

Laubach, T., & Williams, J. C. (2003). Measuring the natural rate of interest. *The Review of Economics and Statistics, 85*(4), 1063–1070.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy, 1*, 19–46.

Mankiw, N. (2007). *Macroeconomics* (6th ed.). Worth Publishers.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science, 4*(1), 141.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29*(2), 449.

Mundell, R. A. (1961). A theory of optimum currency areas. *The American Economic Review, 51*(4), 657–665.

Prasad, E. (2015). *The dollar trap: How the U.S. dollar tightened its grip on global finance*. Princeton University Press.

Rebonato, R. (1999). *Volatility and correlation*. Wiley.

Samuelson, P. (1964). Theoretical notes on trade problems. *Review of Economics and Statistics, 46*, 145–154.

Shambaugh, J. C. (2012). The Euro's three crises. *Brooking's Papers on Economic Activity, 43*, 157–231.

Shreve, S. E. (2000). *Stochastic calculus for finance II: Continuous-time models*. Springer.

Taylor, J. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*, 195–214.

Triffin, R. (1960). *Gold and the dollar crisis*. Yale University Press.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

Index

A. S. Iqbal, *Foreign Exchange*,
https://doi.org/10.1007/978-3-030-93555-9