CHAPTER

10 MONEY AND BANKING

Chapter Outline

Learning Objectives

- Gain an understanding of definitions of the money supply and financial assets.
- Recognise the sharp distinction between the MMT and orthodox representations of the process of credit creation by banks.
- Be able to interpret a bank balance sheet and incorporate changes via flows of new transactions.

10.1 Introduction

In this chapter, we have several objectives. First we will introduce students to commonly used definitions of the **money supply**. Frequent reference has been made in earlier chapters to the purchase or sale of financial assets by both the government through the central bank and treasury, as well as by banks. Here we will provide students with a clear understanding of the generic characteristics of these **financial assets**. We then devote space to developing an understanding as to how banks behave in a modern monetary economy. In the process, we will expose some long-standing myths about the role that banks play in the operation of the financial system.

10.2 Some Definitions

Monetary aggregates

Economists and commentators draw inferences about the economy from trends over time in monetary aggregates. Several measures of monetary aggregates have been devised over the years, but there is some variation across countries in what components are included under each measure. The different measures published by central banks are sometimes summarised as *M0*, *M1*, *M2*, *M3* and *M4* and reflect varying degrees of liquidity (convertibility into cash). It is common to consider the highest-liquidity measures M0 and M1 as **narrow money** while M2, M3 and M4 are considered to measure **broad money**.

M0 is also termed the **monetary base**. In countries such as Australia and the UK it includes circulating notes and coins held by the non-government sector, including banks; the deposits of banks with the central bank (generally called 'reserves'); and other central bank liabilities to the non-government sector. In the USA the monetary base is defined in the same way, but the term M0 is not used. The monetary base is the most liquid measure of the money supply and is also sometimes referred to as **high powered money** (HPM), due to its use as a reserve that is leveraged by private banks that issue their own money-denominated liabilities such as deposits.

M1 typically comprises notes and coins in circulation plus current bank deposits held by the private non-bank sector. In some nations, it includes travellers' cheques and deposit accounts that cheques can be written against. It is also a liquid measure of the money supply because its components are readily available to be used for spending on goods and services.

The US Federal Reserve defines M2 as M1 plus most savings accounts, money market accounts, retail money market mutual funds, and small denomination time deposits (certificates of deposit of under $100,000). M2 is a less liquid measure of the money supply and movements in it are typically used to forecast inflation.

M3 broadens the narrow measures to include less liquid components such as long-term time deposits. Even broader still are M4 measures which add other illiquid assets to the aggregate, such as borrowings from the private sector by non-bank financial intermediaries.

Not all measures are published by all central banks. The US, for example, only publishes the monetary base, M1 and M2. In the UK, there are only two official money supply measures (M0 and M4). The European Central Bank publishes M1, M2 and M3, while in Australia the central bank publishes M0, M1, M3 and M4 (or broad money measure).

10.3 Financial Assets

If a household engages in saving (a flow per period of time) over a number of months or years, then it will accumulate a growing stock of wealth. The household needs to decide whether to continue to add its saving to its existing deposits at its bank or put together a portfolio of financial assets which have different degrees of risk, for example, stocks (shares) or bonds, and are also denominated in the money of account.

Treasuries in modern economies issue bonds which are debt of various maturities; also called securities. These financial assets are bought and sold by the central bank, private banks and the private sector. Private entities (for example, corporations) also issue bonds.

In general, the bond acknowledges that the **issuer** is indebted to the **bondholder**. The bond issuer must pay interest to the bondholder on a periodic basis, and repay the principal (face value of the bond) when the bond matures. Bonds represent wealth for bondholders.

Thus, a **bond** is a **formal contract** to repay a loan (IOU) with interest at fixed intervals. The bondholder is the lender (creditor). The borrower (debtor) issues the bond and the **coupon** is the interest rate paid on the **face value** of the bond, and usually printed on the bond. In these cases, the periodic interest payments are constant.

The **issue price** is what investors pay for the bond when it is first issued. Later, bonds may be traded; at a premium (above par, if good quality, so that there is minimal default risk by the issuer) or at a discount (price below par). The bonds of a currency-issuing government carry zero default risk because such a government can always meet its outstanding liabilities. For this reason, these bonds are very desirable in times of uncertainty.

A **consol** is a special type of bond called a perpetuity, which means there is no maturity date. Interest is paid on this asset forever.

When we talk of the **government bond market** we need to differentiate between the primary and secondary bond markets. A **primary market** is the institutional machinery by which the government sells debt to the non-government sector. While many mistakenly believe that the issuance of bonds in the primary market is designed to raise funds to facilitate government spending, the reality is that currency-issuing governments are not financially constrained (see, for example, Chapter 1) and therefore we must seek a different explanation of why such governments issue debt at all to the non-government sector. We deal with those questions in more depth in Part E of this textbook (*Economic Policy in an Open Economy*).

A **secondary market** is where existing government bonds are bought and sold by interested parties after the bonds enter the monetary system via the primary market. The same arrangements apply, for example, to private share issues (also called equities or stocks). The company raises funds via the primary issuance process and then its shares are traded in secondary markets.

Government bonds are thus negotiable because ownership of the certificate can be transferred (sold) to another owner in the secondary bond market. However, it is important to understand that once the bond is issued, subsequent trading has no impact at all on the volume of financial assets in the system since it just shuffles this wealth between wealth holders.

The process of issuance in primary markets varies across nations and has also varied over time. A typical arrangement in the past was that government bonds would be sold to selected dealers (for example, banks) on a periodic basis in the primary market. Government would determine how much debt it wanted to issue (expressed in the money unit of account) and set a yield it was prepared to pay to the purchasers. The terms offered by this 'take it or leave it' approach might not be attractive to the non-government sector at the time of offer so any shortfalls in purchases of what the government desired to issue would be taken up by central banks. This is a case of government issuing debt to itself, which brings into question the whole logic of issuing debt.

In the late 1970s, the dominant school of economic thought was Monetarism, which erroneously claimed that central bank purchases of debt would be inflationary. Governments fell prey to that logic and started to devise ways to preclude their central banks from purchasing unsold debt. Governments would thus set yields and sell as much debt as possible, but would continuously adjust the yields up or down to meet the market requirements and ensure that there were no discrepancies between net spending (fiscal deficits) and bond sale revenue.

This system gave way to a purer auction system which avoided any claims that the government was manipulating yields, again in response to calls for more 'free' market activity. These auction or tender systems became dominant internationally. In general terms, the treasury would announce the terms of the auction, including how much debt was available for sale, the maturity dates of the debt, and the coupon rate (the periodic interest to be paid on the face value of the bond). The issue would then be put out for tender and then the bond dealers in the primary market would determine the final price of the bonds issued – thus taking discretion away from the elected government in terms of the yields that it would pay on government debt issuance.

As an example, imagine a $1,000 bond had a coupon of five per cent, meaning that you would get $50 per annum until the bond matured, at which time you would get $1,000 back. At the time of issue, the bond market dealers desired a yield of six per cent to satisfy their profit expectations. In this case, the initial specification is unattractive. Prior to the adoption of an auction system, private bond dealers would avoid purchasing the bond under such conditions. But under the auction system they could put in a purchase bid lower than the $1,000 to ensure they got the six per cent return sought on the price that they were willing to pay.

It is important to understand that there is an **inverse relationship between the traded price of a fixed income bond and its yield (rate of interest)**. Why is that so? The general rule for fixed income bonds is that when their prices rise in secondary markets, the yield falls and vice versa. This is because if one pays more to purchase a bond, the coupon payments represent a lower return on the purchase price; on the other hand if one pays less, then the coupon payments represent a higher return. Furthermore, the price of a bond can change in the marketplace according to interest rate fluctuations, even though the bondholder will still only get the face value of the bond back upon maturity.

When interest rates rise elsewhere in the economy, the price of previously issued bonds falls because they are less attractive in comparison to the newly issued bonds, which are offering higher coupon rates (reflecting current interest rates). When interest rates fall, the price of older bonds increases as they become more attractive given that newly issued bonds offer a lower coupon rate than the older higher coupon rated bonds.

The government department that manages these auction processes receives tenders from the bond market traders in the primary bond market. These will be ranked in terms of price (and implied yields desired) and the quantity requested in dollar terms. The bonds are then issued in order of the highest price bid down until the volume the government desires to sell is achieved. So, the first bidder with the highest price (lowest yield) gets their desired volume (as long as it doesn't exhaust the whole tender, which is unlikely). Then the second bidder (higher yield) receives their allocation and so on. In this way, if demand for the tender is low, the final yields will be higher and vice versa.

Bonds are also issued and sold in primary markets by state or provincial governments, multinational and local companies, credit institutions and other public bodies. Companies can finance new capital investment by one or more of the following: (i) issuing bonds, (ii) using retained profits, and (iii) launching a new share issue.

Treasuries and other institutions issue bonds with different times to maturity. For example, the US Department of the Treasury issues bonds of one month, three month, six month, one year, two year, three year, five year, seven year, ten year, twenty year and thirty year duration; a ten year treasury bond matures in ten years, and so on.

Yield concepts in fixed income investments

The yield indicates the return that will be returned from the investment and is usually expressed in percentage terms. There are several concepts of yield that are used in the markets.

- **Coupon or Nominal Yield** – If a bond has a face value of $1,000 and is paying eight per cent in interest (the coupon rate), then the nominal yield is eight per cent. The investor will thus receive $80 per annum until maturity. The coupon yield remains constant throughout the life of the bond.
- **Current Yield** – Suppose you purchase an eight per cent $1,000 bond for $800 in the secondary market. Irrespective of the price you pay, the bond entitles you to receive $80 per year in coupon payments. But unlike the previous example, the $80 payment per year until maturity represents a higher current yield than eight per cent on your investment because it is based on your purchase price of $800. The actual yield is $80/$800 = ten per cent. So, to compute current yield you simply divide the coupon by the price you paid for the bond. In general, if you buy the bond at a discount to face value, the current yield will be greater than the coupon yield, and if you buy at a premium then the current yield will be below the coupon yield.
- **Yield to Maturity (YTM)** – The current yield does not consider the difference between the purchase price of the bond and the principal payment at maturity. YTM considers that in addition to earning interest, an investor can make a realised capital gain or loss by holding the bond until its maturity date. YTM is a measure of the investor's true gain over the life of the bond and is the most accurate method of comparing bonds with different maturity dates and coupon values.

BOX 10.1 WORKED YIELD EXAMPLE

Assume you pay $800 for a $1,000 face value bond in the secondary market. The $200 discount on the face value is considered income or yield and must be included in the yield calculations. Assume that the eight per cent $1,000 bond has five years left to maturity when bought for $800.

A comparison of three yield concepts gives:

- Coupon yield of eight per cent ($80 income flow divided by $1,000 face value).
- Current yield of ten per cent ($80 income flow divided by $800 discounted purchase price).
- YTM of 13.3 per cent. The working is given in the main text below.

The computation of YTM is complex and can be simplified to the following rule of thumb:

$$YTM = (C + PD)/[0.5 \times (FV + P)]$$

where *C* is the coupon, *PD* is the prorated discount, *FV* is the face value, and *P* is the purchase price. *PD* is the difference between the face value and the purchase price, divided by the number of years to maturity. If the bond is trading at a premium, the *PD* is negative which means that the *YTM* is less than the coupon yield.

Using the data in the Worked Example, therefore:

$$YTM = [80 + (200/5)]/[0.5 \times (\$1{,}000 + \$800)] = \$120/\$900 = 13.3 \text{ per cent.}$$

When bond traders talk about yield they are usually referring to the YTM measure which is the only measure that takes into account the effect of principal price, coupon rate, and time to maturity of a bond's actual yield.

There are two ways we can use data on yields for government bonds of different maturities to assess the state of the economy and the degree to which the non-government sector expects inflation to increase in the future. We have seen that rising yields signal weakening prices due to falling private demand for the assets in question. This could reflect a strengthening economy with investors being prepared to acquire more risky assets and less very safe ones. This is also usually when the central bank pushes up the target interbank rate and bond yields more or less follow (see Chapter 20). Further, we can use the movements in yields to gauge what is happening to inflationary expectations in the non-government sector. Rising yields on longer-term maturities indicate that the private markets sense inflation will rise in the future and so they desire to protect real yields by increasing the nominal yields on the bonds.

The second way of looking at the yields is to consider the **yield curve**. The yield curve is a graphical depiction of the **term structure of risk-free interest rates** and plots the maturities of different government bond on the horizontal axis against their respective yields (rates of return) on the vertical axis (see Figure 10.1, which shows the US Treasury yield curve for 3 February 2016).

There are various theories about the yield curve and its dynamics. All share some common notions, in particular that the higher is expected inflation, the steeper the yield curve will be, other factors being equal.

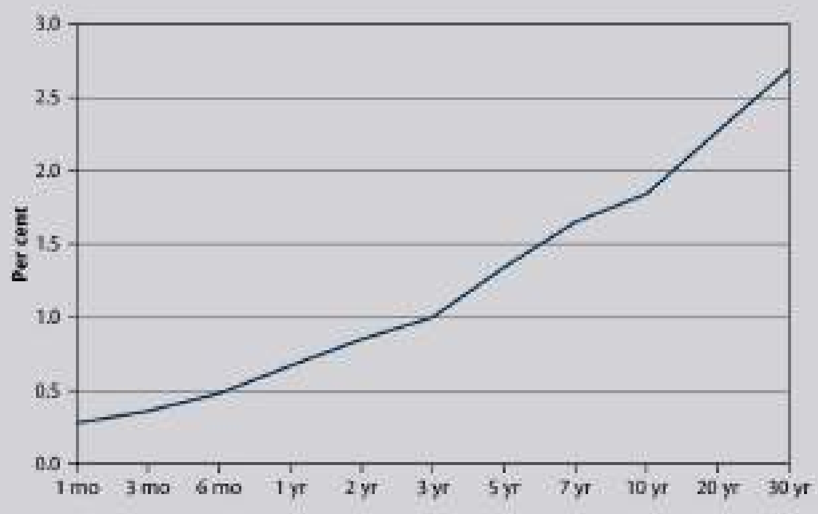
The basic principle linking the shape of the yield curve to the economy's prospects is explained as follows. The short end of the yield curve reflects the interest rate set by the central bank, which sets the competitive rate for cash (highly liquid assets) in the economy. As the short-term interest rate rises (falls), the yields on other less liquid assets will follow suit. The steepness of the yield curve then depends on the yield of the longer-term bonds, which are set by the market. But the short end of the curve is the primary determinant of its slope. In other words, the curve steepens mainly because the central bank is lowering the official cash rate, and it flattens mainly because the central bank is raising the official cash rate.

Bond traders link the dynamics of the yield curve to their expectations of the future economic prospects that are expected to influence central bank interest rate policy. It must be remembered that if central banks raise interest rates, then this will tend to cause prices of bonds to fall. This is called a **capital loss**. The prices of bonds with the greatest term to maturity will tend to be affected the most, so longer-term bonds are generally subject to the greatest risk of capital loss. For this reason, there may be a link between inflation expectations, expectations of central bank policy, and prices and yields on longer-term bonds.[1]

In summary, there are three shapes that the yield curve can take:

- **Normal** – Under normal circumstances, short-term bond rates are lower than long-term rates. The central bank attempts to keep short rates down to keep levels of activity as high as possible and bond investors desire premiums in longer-term maturities to protect them against capital losses. Thus, the yield curve is upward sloping, as in the case shown for the US in Figure 10.1.
- **Inverted** – Sometimes, short-term rates are higher than long-term rates and the yield curve is said to be inverted. When the economy starts to overheat, the expectations of rising inflation that might induce the central bank to raise its target interest rate lead to higher bond yields being demanded on assets with longer-term

Figure 10.1 US Treasury yield curve (3 February 2016)



Source: Authors' own. Data from Daily Treasury Yield Curve Rates, Resource Center, US Department of the Treasury (https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/default.aspx).

maturities. The central bank might respond to the building inflationary pressures by raising short-term interest rates sharply. Although bond yields might rise, the significant tightening of monetary policy causes short-term interest rates to rise faster, resulting in an inversion of the yield curve. The higher interest rates may then lead to slower economic growth.

- **Flat** – A flat yield curve is seen most frequently in the transition from positive to inverted and vice versa. As the yield curve flattens, the yield spreads drop considerably. A yield spread is the difference between say, the yield on a one year and a ten year bond. What does this signal about the future performance of the economy? A flat yield curve can reflect a tightening monetary policy (short-term rates rise). Alternatively, it might depict a monetary easing after a recession (easing short-term rates) so the inverted yield curve will flatten out.

Movements in the yield curve are thus closely watched by economists due to the information that they convey about the general health of the economy, possible central bank interest rate adjustments, and inflationary expectations in the non-government sector.

BOX 10.2 THE ORTHODOX APPROACH TO NOMINAL INTEREST RATE DETERMINATION: THE FISHER EFFECT

One risk in holding a fixed coupon bond with a fixed redemption value is purchasing power risk.

Orthodox economists who adopt the loanable funds approach to interest rates believe that most people would prefer to consume now rather than later. To encourage forgone consumption now, a yield on savings must be provided by markets. The yield is intended to allow a person to consume more in the future than has been sacrificed now. But if the prices of goods and services increase in the meantime, then inflation could completely wipe out any gain in real consumption, so that the real interest rate is zero.

Consider a person who invests in a one year $1,000 coupon treasury bond with an expected single coupon payment of $100. The individual will expect to get $1,100 on the redemption date.

Assume that over the holding period, prices rise by ten per cent. At the end of the year, a basket of goods that previously cost $1,000 would now cost $1,100. In other words, the investor is no better off at the end of the year as a result of the investment. The nominal yield has been offset by the price inflation. Orthodox economists believe that investors are motivated by 'real' returns, not by nominal returns. This is because they view the decision to invest as coming from consumers who choose whether to consume now or to consume in the future, with consumption taking the form of real goods and services. If such savers do not take account of inflation, their future real consumption will be less than they desired.

Orthodox economists propose that the nominal interest rate must equal a real interest rate plus expected inflation. The real interest rate is supposed to be the market-determined real return that will equate the saving supply of funds with the investment demand for funds. It is thus a real equilibrium interest rate. However, as contracts are written in nominal terms, that is, in terms of nominal interest rates, the nominal rate must include compensation for the expected inflation rate. This addition to the real rate as inflation expectations rise is called the Fisher effect, named after the American economist Irving Fisher, who identified this relationship in the 1930s. Many market participants believe this applies to bond markets, and there is a strong belief that nominal yields are adjusted by markets to preserve purchasing power.

Purchasing power risk increases as the maturity lengthens. This is one reason many economists believe that longer maturity rates will generally be higher. The market yield is equal to the real rate of return required plus compensation for the expected rate of inflation. If the inflation rate is expected to rise, then market rates will rise to compensate. In this case, we would expect the yield curve to steepen, given that the Fisher effect will impact more significantly on longer maturity bonds than at the short end of the yield curve.

10.4 What Do Banks Do?

The neoclassical view: the money multiplier

In most textbooks, banks are presented as financial intermediaries that take in deposits, hold a small fraction of these in the form of reserves, and lend out the remainder. The causality is from deposits to reserves to loans. If each bank follows these principles in making loans, aggregate lending expands through the 'deposit or money multiplier'. For the moment assume that all banks are required to maintain a ratio of reserves to deposits of ten per cent. This might be to enable them to readily respond to a loss of reserves resulting from spending by customers (on goods and services, say) whose sellers bank elsewhere, or customers seeking to hold additional cash.

This is how the neoclassical school of thought describes the operation of the **money multiplier**:

i) Assume that a customer deposits \$100 in Bank A;
ii) Bank A retains \$10 of reserves to conform to the required reserves-to-deposit ratio of ten per cent. To expand its loan portfolio and increase profits, the remaining \$90 is loaned to a customer whose deposits rise by \$90;
iii) The customer spends these deposits and the recipient of the funds (seller) deposits \$90 in Bank B;
iv) Bank B then lends $0.9 \times \$90 = \$81$ (keeping ten per cent, that is, \$9 as additional reserves, as required) to a customer to finance their expenditure and so on.

At each stage the amount lent and then spent diminishes. It can be readily shown that if this was the way the banking system operated, then \$900 of additional loans are created. With the initial new deposit, this means that deposits have risen by a total of \$1,000 and are 'backed' by \$100 of reserves, thereby conforming to the required ten per cent ratio.

This example is what the mainstream textbooks call a **fractional reserve banking system**. It purports to explain how banks create money, which increases the *M*1 money supply due to the increase in current deposits. In terms of the initial deposit of \$100, the multiplier is 10, which is the inverse of the required reserves-to-deposit ratio of 0.1. A smaller money multiplier results if the non-government sector chooses to hold more cash when credit is created.

Note that no individual bank 'creates money' in this example, but the system as a whole 'multiplies' the initial deposit of $100 into $1,000. At each step, each bank simply lends out 90 per cent of the deposit it has received, keeping ten per cent as reserves. According to mainstream textbooks, the 'magic' results from fractional reserve banking. The larger the fraction of a deposit that must be retained as reserves, the smaller the multiplier effect. Following this logic, if the reserve ratio were zero (no reserves held against deposits), the banks would create an infinite amount of deposits after the deposit of just one dollar.

The standard textbook example is typically assumed as a ten per cent ratio, so that students can readily calculate a money multiplier equal to ten! On 12 April 1992, the US Federal Reserve Bank, for the first time in history, set the required reserve ratio on demand deposits at the magical ten per cent, making theory appear to coincide with reality. But that coincidence did not make the theory correct. As we will see next, the money multiplier as a description of modern banking is a myth, and bears no relation to how banks operate in the real world.

To summarise the dominant neoclassical view, banks are conceived as being financial intermediaries that maximise profits. They take in deposits to build up reserves so that they can then on lend the deposits at a higher interest rate. However prudential regulations require that they maintain a minimum reserve-to-deposit ratio. The fractional reserve requirements mean that the resulting credit creation process is finite.

In addition, many economists still believe that the monetary base, which consists of bank reserves and cash held by the non-government sector, is under the control of the central bank. Thus, by controlling the size of the monetary base and setting the required reserve ratio, the central bank is considered to be able to control the size of the money supply or the quantity of money.

Thus in the neoclassical narrative, the money supply is considered **exogenous** and determined by the central bank. This is an important claim because it has underpinned arguments that central banks can cause inflation by allowing the money supply to grow too quickly. From this follows the Quantity Theory of Money's (QTM) policy recommendation that the central bank can fight inflation by slowing money growth. As we will see in Chapter 20 (and analyse in Chapter 23), the QTM is a flawed conception of the inflation generation process. We will also demonstrate that the central bank does not have the capacity to control the money supply in a normally functioning money system.

The implication of the operation of the money multiplier is that a bank would forgo profitable loan opportunities if it did not have sufficient reserves to enable additional credit creation. Some allowance is made for discretion: the deposit multiplier is claimed to be a function of interest rates and interest rate differentials, bank preferences regarding their holdings of excess reserves, and also public preferences regarding their holdings of cash, as noted, and time deposit and demand deposit ratios. However, as Brunner (1968) 'demonstrated', these factors are of only minor importance.

MMT representation of the credit creation process

The neoclassical characterisation of the credit creation process, which is driven by fractional reserve requirements, is not an accurate depiction of the way banks operate in a modern monetary economy characterised by a fiat currency and a flexible exchange rate.

In the real world, the business of banking is complicated but is, in some respects, similar to that of other profit-seeking firms. Like these other firms, banks seek to earn profits and thereby generate returns for shareholders. Making loans secures profits, as long as the banks are paying a lower rate of interest on the funds that they borrow than they receive from their customers who take out loans.

First, a necessary condition for credit creation is that there are non-bank firms and/or households who are seeking loans to finance their planned spending on goods, services or assets. Second, some of these entities must be considered creditworthy by the banks, so that there is a high probability that the loan will be repaid in full. What constitutes creditworthiness varies over the business cycle and lending standards tend to become more lax in boom times as banks chase market share. Third, the banks must anticipate that there is profit to be made by making these loans, as described above.

Banks make loans independently of their reserve positions (that is, their holdings of reserves, relative to their liabilities). After originating loans they will borrow additional reserves if required by law or for clearing purposes.

Bank managers generally neither know, nor care, about the aggregate level of reserves in the banking system. Certainly, no loan officer ever checks the individual bank's reserve position before approving a loan. Bank lending decisions are affected by the price of reserves and expected returns, not by reserve positions. If the spread between the rate of return on an asset (a security or a loan) and the cost of borrowing reserves is wide enough, even a bank that is already deficient in reserves will purchase the asset or make a loan and cover the reserves needed by purchasing (borrowing) reserves in the **interbank market**. The interbank market connects the banks which lend reserves to and borrow reserves from each other when needed.

The important point is that when a bank originates a loan to a firm or a household, it is not lending reserves. Bank lending is not easier if there are more reserves, just as it is not harder if there are less. Bank reserves do not fund money creation in the way that is claimed in the money multiplier and fractional reserve deposit story; banks do not wait for deposits to come in before they make loans.

The main difference between banks and other types of firms involves the nature of the liabilities. Banks 'make loans' by purchasing the IOUs of 'borrowers'. This results in a bank liability, usually a demand deposit, at least initially, that shows up as an asset of the borrower. Thus, a customer of a bank who secures a loan is simultaneously a 'creditor' of the bank, due to holding a demand deposit, but also a 'debtor' to the bank. These creditors will almost immediately exercise their right to use the newly created demand deposits as a medium of exchange for purchases of goods and services, or assets. Bank liabilities (bank deposits) are used by households and non-bank firms for transactions in the form of cheques or transfers. Customers can also redeem demand deposits at par (dollar for dollar) against fiat money (which is guaranteed by the government) to enable cash to be used for purchases or making payments that are due. The government will also accept some kinds of bank liabilities in payment of taxes.

In turn, bank reserves are used for payment (or interbank settlement) among banks and for payments made to the central bank. Thus, when bank 'creditors' draw down their demand deposits, by either spending or choosing to hold more cash, a corresponding loss of reserves for the individual bank results. The bank may then either sell an asset or increase its liabilities (borrowing additional reserves) to cover the loss of reserves.

The interbank market (called the federal funds market in the US) functions to shuffle the reserve balances that the member (private) banks keep with the central bank to ensure that each of these banks can meet its reserve targets, which might be simply zero balances at the end of a specified period of time (that for simplicity, we could assume is a day). In aggregate, however, such activities only shift reserves from one bank to another. If more reserves are needed in total, they must be supplied by the central bank.

Far from waiting for deposits before they create loans, banks in the real world expand their balance sheets by lending as described below.

Loans create deposits

Loans create deposits that are then backed by reserves after the fact. The process of extending loans (credit), which creates new bank liabilities, is unrelated to the reserve position of the bank. In the pursuit of profit, banks take applications from creditworthy customers who seek loans and assess them according to the verity of the application, although in the lead-up to the Global Financial Crisis of 2008, the validation process became very lax.

The only thing that constrains the bank loan desks from expanding credit is a lack of creditworthy applicants, which can be due to banks raising the qualifying standards in times of pessimism, or can occur if creditworthy customers are loath to seek loans because of future uncertainty.

The major insight is that any balance sheet expansion that leaves a bank short of the required reserves may affect the return it can expect on the loan. This is a consequence of the 'penalty' rate the central bank might exact through the discount window (the central bank facility for lending to banks in need of reserves) should the bank fall short of the reserves it requires at the end of the day to cover the claims on it. However, it will never impede the bank's capacity to make the loan in the first place. It is thus quite wrong to assume that the central bank can influence the capacity of banks to expand credit by adding more reserves into the system. We will address this proposition in more detail in Chapter 23.

Banks do not loan out reserves

A corollary of the 'loans create deposits' insight is that banks do not loan out reserves, which raises the question of what role do bank reserves actually play?

Banks must hold reserve balances with the central bank as part of the **payments system**. The reserves are used to make interbank payments. Each day millions of transactions are reconciled (settled) through these interbank payments. For example, cheques drawn on Bank A and deposited at Bank B will see funds transferred from Bank A's reserve balances to those of Bank B.

If a particular bank finds itself short of the quantity of reserves necessary to resolve all the daily claims against it, then it can first try to borrow reserves from other banks that might have excess reserves in relation to their requirements on that particular day. But, as we will see in Chapters 20 and 23, an overall shortage (or excess) of reserves across the banking system must be rectified by the central bank which provides reserves to banks (in the case of a shortage) or may drain them from the system in the case of an excess. This central bank intervention is what we refer to as its liquidity management role and allows the bank to manage the overall level of reserves so that it is consistent with its interest rate target. For example, if on any particular day there is an excess of reserves (over and above the quantity required to settle transactions) and the central bank does not offer any competitive return on them, banks holding those excesses will try to loan them out overnight, which has the effect of driving down the short-term interest rate. The central bank must drain those reserves (by selling government bonds to the banks in return for debits to the reserve accounts) to ensure the overnight interbank interest rate remains equal to its desired policy (target) rate. We will learn more about this in Chapter 23.

Endogenous money

We have stated that unlike the story presented in neoclassical textbooks, in the real world the central bank cannot control the money supply. In other words, the money supply is **endogenous money** in the sense that the supply of bank money is determined 'endogenously' by the demand for bank loans, plus the willingness of banks to lend (which gives rise to the creation of deposits). The neoclassical theory erroneously believes that the money supply is exogenous and determined through the money multiplier interacting with the monetary base, which neoclassicists believe to be under the control of the central bank.

The demand for bank loans is determined by the spending decisions of private economic agents (including decisions regarding asset purchases). These can be affected, but only very indirectly, by the loan rate of interest. Banks supply loans only because someone is willing to 'borrow' bank money by issuing an IOU to banks. This means that the interest rate cannot be determined by the supply of and demand for loans since supply and demand are not independent. Rather, banks are price setters in short-term retail loan markets. They then meet the demand for loans with some quantity rationing at that price. In other words, some requests for loans are refused, even where aspiring borrowers claim to be willing and able to pay the going interest rate.

There can be several reasons for such quantity rationing of large segments of the population. Banks might worry about the default risk of some borrowers but might not be able to raise interest rates sufficiently to cover the default risk. Quantity rationing is then superior to price rationing, that is, raising the interest rate charged to some borrowers. Also, banks probably have better information than do borrowers about default risks. For example, the borrower who wishes to open a new restaurant might not have accurate information about bankruptcy rates in the industry or might simply be overly optimistic. On the other hand, banks can never know the future, so must operate based on rules of thumb (for example, informal rules that restrict loan size). Some quantity rationing can even be irrational, perhaps discriminatory, because banks have traditionally forgone certain kinds of loans or are reluctant to lend to certain groups in the community. The key point is that the supply of loans does not simply adjust to the demand for loans at some interest rate.

The short-term retail interest rates can be taken as a mark-up over short-term wholesale interest rates. Exactly what determines the mark-up (and whether it is variable) is controversial, but not important to our analysis here (see Moore, 1988).

Wholesale interest rates, finally, are under the influence of central bank policy. Individual banks use wholesale markets to rectify a mismatch between retail loans and deposits. Most banks will not be able to exactly reconcile their retail loans and deposits. Some banks will be able to make more retail loans than they can retain in deposits and thus suffer a loss of reserves, while others will find fewer loan customers than depositors, so they will have a surplus of reserves. Banks then use wholesale markets to either 'purchase' reserves by issuing wholesale liabilities (for example, negotiable, large denomination certificates of deposit (CDs) or by borrowing central bank funds), while surplus banks will sell their excess reserves.

As discussed above, the central bank sets the overnight interbank rate. This rate then determines other short-term wholesale rates (mainly by marking up, but also by marking down) through arbitrage.

Summary

The neoclassical position is that banks leverage (create credit) when provided with new deposits, but are constrained by fractional reserve requirements. Since the central bank is supposedly able to control the monetary base, it is claimed that the central bank can control the supply of money.

Reflecting what happens in the real world, MMT demonstrates that the central bank cannot control the monetary base, because monetary policy is conducted by the central bank setting a target interbank rate and providing the right level of reserves to the banking system so that banks lend to and borrow from each other at this target rate (for more details, see Chapters 20 and 23). Second, a bank is not constrained by its reserve position in deciding whether to make a loan to a particular customer. If the customer appears creditworthy and the loan is profitable to the bank, it will make the loan and then acquire sufficient reserves by borrowing from other banks or the central bank. Thus, in contrast to the neoclassical position of deposits driving loans, MMT shows that loans drive deposits. Third, taken together, the growth in the broad money supply is driven by the demand for loans and the monetary base adjusts to the pressures that the endogenous monetary growth places on the central bank in its quest to sustain a particular policy interest rate. Hence the supply of money is determined endogenously while the price of money (short-term interest rate) is determined exogenously by central bank policy.

An example of a bank's credit creation: a balance sheet analysis

The balance sheet of a typical bank looks like that in Figure 10.2.

The entries on the balance sheet are the cheque and savings accounts. Note that they are the IOUs of banks, and hence appear as liabilities. The bank promises to convert deposits in a cheque account (and deposits in most savings accounts) into cash on demand. Banks hold financial assets in the form of loans to customers and securities (that is, treasury debt and other financial assets).

Firms in general and banks should have positive net worth, which is the difference between total assets and total liabilities. Total Assets in the left-hand column will balance with the items in the Liability column, because the latter includes net worth.

The following simplified series of balance sheets will clarify the process of credit creation by Bank A. Let us assume that Bank A starts with the very simple balance sheet in Figure 10.3, which is expressed in terms of stocks.

Figure 10.2 A typical bank balance sheet

Assets	Liabilities
Advances (loans) Securities Reserves Other assets	Cheque accounts Savings accounts Other liabilities **Net worth**

Figure 10.3 Bank A initial balance sheet

Assets		Liabilities	
Building	$200		
		Net worth	$200

Its owners have raised capital and bought the building. The owner's equity or net worth is equal to the value of the building that they have purchased. Bank A has not engaged in any banking activity yet.

Now a customer comes into the bank and says that they would like to borrow $200 to finance the purchase of a car. The bank checks their creditworthiness by asking for income tax returns, proof of assets, credit history, and so on. If the customer is approved, then the bank's balance sheet takes the form shown in Figure 10.4.

The bank just created $200 of money entries (deposits in the cheque account of the customer in return for the customer's IOU, or promise to pay $200). The bank's total assets, liabilities plus net worth, are now $400.

Before we move on to the customer's spending of their deposit, let us examine this balance sheet carefully.

Where did the bank get the money entry it created?

- A cheque account was created *ex nihilo*, that is, from nothing, by entering a number (200) in a computer ledger on behalf of the borrower. In the past, banks could also issue their own banknotes, but generally only central banks can do that now.
- The bank did not need any prior deposits, or any cash in its vault. In fact, the bank did not have any cash in its vault, nor did it have any deposits in its account at the central bank in this simplified example.
- The bank is not lending anything it has, it just creates money entries (that is, bank deposits), at will.
- Those money deposits or entries are its liabilities/IOUs.
- By creating those bank IOUs, the bank promises to:
  - Convert deposits into cash on demand;
  - Accept any of those IOUs in payment of debts owed to the bank.

The cheque account is just a legal promise to convert to cash on demand, and to accept payment in the form of the bank's own IOUs. The bank does not have to have any cash now.

The success of the banking operation (lending by accepting an IOU, and the creation of a demand deposit) depends on:

- The capacity of the customer to repay, that is, their creditworthiness. If they have problems in making timely payments on their debts, this affects the value of the bank's assets and its own income inflows and ultimately affects the net worth of the bank, the bank's capital ratio, and the shareholders' return on equity.
- The bank's capacity to acquire reserves at low cost if:
  - The customer wants to withdraw cash;
  - The bank needs to pay debts to other banks through an interbank settlement following the customer's spending (see below);
  - The bank needs to settle tax payments made by the customer to the government.

Figure 10.4 Bank A balance sheet showing loan

Assets		Liabilities	
Loan to customer	+ $200	Cheque account of customer	+ $200
Building	$200		
		Net worth	$200

If these conditions are not satisfied the bank gets into trouble; it can become insolvent or illiquid. Insolvency means that the bank's net worth falls to or below zero; illiquidity means that it cannot meet cash withdrawals or clearing. Thus, even though banks can create unlimited amounts of money

Figure 10.5 Bank A balance sheet showing purchase of car

Change in Assets	Change in Liabilities
	Cheque account of the customer −$200 Reserves due to Bank B +$200

Figure 10.6 Bank B balance sheet showing purchase of car

Change in Assets	Change in Liabilities
Claim on Bank A reserves +$200	Cheque account of car dealer +$200

deposits, they have no incentive to do so because they may become unprofitable.

What happens if the customer now pays $200 to a car dealer who has a bank account at Bank B? The balance sheets of Banks A and B look like Figures 10.5 and 10.6, respectively. (Note that we are now just dealing with the change in assets and liabilities rather than their levels.)

Bank A's liabilities in the form of the customer's cheque account have dropped by $200 through the purchase of a car, but the transaction is not confined to the reduced balances in the customer's account at Bank A and the increased balances of the car dealer at Bank B. Bank A now owes Bank B $200 and needs reserves to settle this debt, but does not have reserves. Where does it get the reserves?

The banks are required to keep reserve accounts at the central bank. These reserves are liabilities of the central bank and assets of the banks, and function to ensure that the payments (or settlements) system functions smoothly. That system relates to the millions of transactions that occur daily between banks as cheques are tendered by citizens and firms and other bodies. Without a coherent system of reserves, Bank A could easily find itself unable to fund Bank B's demands based on the cheque drawn on the customer's account and presented at Bank B by the car dealer.

Bank A will get the reserves from the source that is the least costly. It may sell assets, but in our example, Bank A only has a building so it would be very costly to get reserves that way. Bank A could sell bonds if it had any, or it could borrow reserves from other banks (domestic or foreign) or the central bank. A common way to get the reserves is to borrow from the central bank, which is the monopoly supplier of reserves. Figure 10.7 documents the latest change to Bank A's balance sheet, associated with obtaining these reserves, while Figure 10.8 shows the changes to the central bank's balance sheet.

Now that Bank A has the reserves it can settle its debt with Bank B. The changes to the two banks' balance sheets are shown in Figures 10.9 and 10.10.

The debt between the two banks has been settled. The final balance sheet of Bank A looks like Figure 10.11.

Bank A makes money as long as the interest it receives on the loan to the customer is higher than the interest it pays to the central bank on the reserves.

Figure 10.7 Bank A balance sheet showing loan from central bank

Change in Assets	Change in Liabilities
Reserve +$200	Debt to central bank +$200

Figure 10.8 Central bank balance sheet showing loan

Change in Assets	Change in Liabilities
Reserve loan to Bank A +$200	Reserve +$200

The balance sheet of Bank B is shown in Figure 10.12. We assume that Bank B had reserves prior to the cheque account of the car dealer being increased by the sale of the car to the customer.

The final balance sheet of the central bank after all transactions is shown in Figure 10.13.

Note that none of these operations involved any transfer of physical cash. It was all bookkeeping entries conducted digitally through computer networks.

Also, note that we have only shown the assets and liabilities directly related

Figure 10.9 Bank A balance sheet showing settlement of debt

Change in Assets		Change in Liabilities	
Reserves	−$200	Reserves due to Bank B	−$200

Figure 10.10 Bank B balance sheet showing settlement of debt

Change in Assets		Change in Liabilities
Claim on Bank A	−$200	
Reserves	+$200	

to our example. Of course, private banks and the central bank have many other assets and liabilities, as well as net worth on their balance sheets.

In practice, the central bank will usually not advance reserves to the bank directly in the form of an unsecured advance; instead it will ask for collateral (usually a treasury security) in exchange and will provide funds for less than the value of the collateral. So, if Bank A has a $300 bond, it surrenders it to the central bank in exchange for reserves. The central bank might only give the bank $285 if the discount rate is five per cent. The discount rate is one way in which the central bank can try to limit credit creation in the economy.

Figure 10.11 Bank A final balance sheet

Assets		Liabilities	
Funds advanced to customer	$200	Debt to central bank	$200
Building	$200		
		Net worth	$200

Figure 10.12 Bank B final balance sheet

Assets		Liabilities	
Reserves	$200	Cheque account of car dealer	$200

Figure 10.13 Central bank final balance sheet

Assets		Liabilities	
Reserve loan to Bank A	$200	Reserves	$200

Conclusion

It is insufficient and misleading to think of modern banks as 'intermediaries' that take in deposits and then lend most of them out, while retaining some fraction as reserves. Instead, we should think of banks as making loans (accepting the IOUs of borrowers) and then creating demand deposits that the borrowers can use to finance their spending. Banks mostly use reserves for clearing, that is, for settling payments with other banks, the central bank, and the treasury, and at the ATM (when cash is withdrawn). Banks obtain reserves as needed either by borrowing them from other banks or through creation of reserves by the central bank. We will explain in more detail how, and why, central banks accommodate the demand for reserves in Chapter 20.

References

Brunner, K. (1968) "The Role of Money and Monetary Policy", *Federal Reserve Bank of St Louis Review*. 50, 8–24.

Moore, B. (1988) *Horizontalists and Verticalists: The Macroeconomics of Credit Money*, Cambridge: Cambridge University Press.

Endnote

1. Orthodox economists propose that the nominal interest rate must equal a real interest rate plus expected inflation. The real interest rate is supposed to be the market-determined real return that will equate the saving supply of funds with the investment demand for funds. It is thus a real equilibrium interest rate. However, as contracts are written in nominal terms, that is, in terms of nominal interest rates, the nominal rate must include compensation for the expected inflation rate. This addition to the real rate as inflation expectations rise is called the Fisher effect (see Box 10.2). Many market participants believe this applies to bond markets. There is a strong belief that nominal yields are adjusted by markets to preserve purchasing power.

Visit the companion website at www.macmillanihe.com/mitchell-macro for additional resources including author videos, an instructor's manual, worked examples, tutorial questions, additional references, the data sets used in constructing various graphs in the text, and more.