

# Money, Inflation, and Output: A Quantity-Theoretic Introduction

Jón Steinsson\*

University of California, Berkeley

July 4, 2021

An amazing variety of items have been used as money throughout history. For many centuries, cowrie shells were a dominant form of money in large parts of Asia and Africa. In Icelandic (my mother tongue), the word for money is the same as the word for sheep (“fé”) because sheep were used as a medium of exchange for centuries. Similarly, the English word “pecuniary” comes from the Latin word “pecus,” which means cattle. Tobacco and beaver skins (among other objects) were used as money in the British colonies of North America in the 17th and 18th centuries. The most common form of money over the past millennium has probably been coins made of gold, silver, copper, and sometimes other metals. More recently, paper money, bank deposits, and other types of ledger entries have become the dominant form of money in much of the world.

Most people are used to the monetary system employed in their own society, and may not give it much thought on a day-to-day basis. But the monetary system used in one society can strike people in other societies as puzzling and even illogical. Take, for example, the Micronesian island of Yap. Its unusual monetary system was described in *The Island of Stone Money* by anthropologist Henry Furness (1910) as well as by Friedman (1992, ch. 1). The medium of exchange on Yap was called fei. Furness describes fei as “large, solid, thick, stone wheels, ranging in diameter from a foot to twelve feet, having in the center a hole varying in size with the diameter of the stone, wherein a pole may be inserted sufficiently large and strong to

---

\*I would like to thank Amanda Awadey for excellent research assistance. I would like to thank Emi Nakamura, Kivanc Karaman, Nuno Palma, Angela Redish, and Alan Taylor for valuable comments and discussions. First posted in June 2021.

bear the weight and facilitate transportation.” Since moving the fei was difficult, they would sometimes be left unperturbed on a previous owner’s premises after a transaction had been concluded. Furthermore, some fei had actually been lost at sea. But this did not matter. One particular family was considered by everyone on Yap to be wealthy on account of an enormous fei, the size of which was only known by tradition since it had been lost at sea several generations earlier. It was universally conceded that the accident of losing the stone to the vagaries of the sea should not affect its marketable value. “The purchasing power of that stone remains, therefore, as valid as if it were leaning visibly against the side of the owner’s house,” Furness explains.

This may strike readers (that are not from Yap) as quite strange. But before passing judgment, consider the practice in other parts of the world of expending enormous effort digging relatively soft metals useless for most purposes (i.e., gold and silver) from deep in the ground, refining them at considerable cost, transporting them long distances, only to bury them again in elaborate vaults deep in the ground. And if that is not strange enough, consider the practice of parting with valuable goods and services in exchange for small rectangular pieces of green paper with pictures of past U.S. presidents. These pieces of paper are not “backed” by anything: one cannot return them to their issuer (the U.S. Federal Reserve) in exchange for anything of “real value.” Why are these particular pieces of paper widely accepted as payment for goods and services in the United States? Can it reasonably be said that the monetary system in the United States today (i.e., in early 2021) is any less mysterious or strange than that of the island of Yap a century ago?

These examples illustrate well that the value of money need not be and rarely is determined by the “fundamental,” i.e., non-monetary, value of the object that is being used as money. This fact makes money and monetary economics rather mysterious. Perhaps even the most mysterious subject in all of economics. The renowned monetary economist Milton Friedman has argued that monetary matters are governed largely by appearance, illusion, or “myth.” Why are very particular (intrinsically useless) pieces of paper highly valued as a medium of exchange? According to Friedman, “the short answer – and the right answer – is that private persons accept these pieces of paper because they are confident others will. The pieces of paper have value because everybody thinks they have value. Everybody thinks they have value because in everybody’s experience they have had value. ... the existence of a common and widely accepted medium of exchange rests on a convention: our whole monetary system owes its existence to the mutual acceptance of what, from

one point of view, is no more than a fiction.” (Friedman, 1992, p. 10)

Monetary economics is the study of this fiction. It seeks to understand how sturdy this fiction is. Under what circumstances does the value of money remain stable or change predictably? What does it take to make this fiction crumble and what happens when it does. Monetary economics also seeks to understand how changes in the value of money affect “real outcomes” in the economy such as the level of output and unemployment. Finally, it is important to understand whether certain forms of money yield better economic outcomes than others. These are the questions that we will be seeking to shed light on in this chapter and the next few chapters.

Monetary economics is intimately related to Keynesian economics and the study of business cycles. In earlier chapters of this book, we have mostly used neoclassical models to analyze various macroeconomic phenomena. Now we transition to mostly use Keynesian models. This chapter introduces a simple version of a Keynesian model, which we will refer to as the medieval economy model. Over the course of the next seven chapters, we will augment and develop this model in various ways until we arrive at a model that we will refer to as the modern business cycle model. I hope that developing our monetary / business cycle model in this step-by-step fashion will both help you understand the role of each piece of the model and also illuminate how different versions of the model can be used productively to understand different phenomena in monetary economics and the study of business cycles. But before turning to these models, we start with a parable.

## **1 A Parable: Baby-Sitting on Capitol Hill**

In the 1970s, Joan and Richard Sweeney were members of a babysitting co-operative in Washington DC called the Capitol Hill Babysitting Co-operative. In a 1977 article, they recounted a sequence of “crises” that befell the co-op (Sweeney and Sweeney, 1977). Although it may seem far fetched a first, the story they tell illustrates well certain core ideas in monetary economics.

The purpose of the Capitol Hill Babysitting Co-op was—as the name suggests—to allow member households to exchange babysitting services. The co-op was quite large with on the order of 150 member households. To keep track of each members’ babysitting balance, the co-op issued scrip. Each piece of scrip was good for one-half hour of babysitting. The scrip was therefore the money in this “babysitting

economy” and the price of goods (babysitting) was constitutionally fixed.

Early on, each household would receive 20 hours worth of scrip when they joined the co-op and were required to pay back 20 hours worth when they left the co-op. Everything worked well for some time. But as time passed, the amount of scrip in circulation gradually fell. Some pieces of scrip were undoubtedly lost. But the co-op also had a built in tendency for the quantity of scrip outstanding to fall over time. The bylaws of the co-op stipulated that each household must pay a certain number of hours worth of scrip in dues each year. These dues were meant to finance co-op expenses, which were mainly payments to “monthly secretaries” who received requests for sitters and tried to fill them (remember, this was before the time of the internet). The problem was that the dues slightly outran the expenses. So, the co-op was taking in slightly more scrip in “taxes” than it was paying out and as a consequence the quantity of scrip outstanding per household was dwindling as time passed.

As the quantity of scrip in circulation fell, each household had less and less. Over time, more and more households started to feel that they didn’t have enough scrip and sought to acquire more. They did this partly by becoming more willing to sit. But they also became more reluctant to spend the few pieces of scrip they had by going out. Their thinking was: What if we really need babysitting at a later date and don’t have enough scrip? Better to conserve the scrip we have and offer to sit for others so as to build up our stock.

But you need two to tango. It doesn’t work for everyone to sit and no one to go out. Since few households wanted to go out, equally few could sit. The babysitting economy had fallen into a recession. Gross Babysitting Product (GBP) was way down and unemployment was rampant: lots of workers wanted to work (babysit) but there were no jobs to be had.

The initial reaction of co-op members (who were mostly lawyers) was to pass a rule mandating that everyone must go out at least ever so often (once every 6 months). As Sweeney and Sweeney describe: “The thinking was that some members were shirking, not going out enough, displaying the antisocial ways and bad morals that were destroying the co-op. Hence the bylaw to correct morals.” But this coercive solution didn’t work.

Finally, in desperation, the co-op resorted to monetary policy: each household was given 10 extra hours worth of scrip. Miraculously, GBP boomed. In Sweeney and Sweeney’s words: “There shortly arrived a balance between those who wanted to go out and those who wanted to sit. A golden age, on a minor scale. Those people

who previously hadn't wanted to go out must have changed their moral—or maybe it was the ten hours all around.”

Unfortunately, the golden age only lasted a couple of years. As time passed, a new problem arose. The opposite of the old problem: more people wanted to go out than were willing to sit. In other words, demand exceeded supply and there were shortages of goods to buy. At the time Sweeney and Sweeney wrote their paper, this new problem had not been solved. But they had a theory as to why it was occurring.

As a part of the monetary reform discussed above, the co-op had changed the rules relating to the amount of scrip households received when they joined and the amount they were required to give back when they left the co-op. After the reform, new households received thirty hours worth of scrip and exiting households were required to pay back only twenty hours worth of scrip. This aspect of the reform implied that each time a household left and a new one joined the co-op, the aggregate amount of scrip in circulation rose by 10 hours. The aggregate amount of scrip in circulation was therefore influenced by the level of turnover of households in the co-op. If turnover was high, the amount of scrip in circulation would rise. If turnover was low, it might fall.

Sweeney and Sweeney pointed out that turnover had been quite high and this had led to a substantial increase in the aggregate quantity of scrip in the babysitting economy. They argued that this was the cause of the imbalance between demand and supply for babysitting services. Over time, each household's holdings of scrip had grown. Eventually many households had so much scrip that they felt no need to babysit.

The travails of the Capitol Hill babysitting economy during the 1970s illustrates well the lesson that an economy with too little money in circulation can suffer a recession due to insufficient demand, while an economy with too much money can suffer from good shortages (too much demand). Somewhere in between is a point one might call the Goldilocks economy, where there is neither too much nor too little money in the economy, rather the amount of money is just right. In this case demand will equal supply and the economy will function efficiently.

But what is it that makes the efficient functioning of the economy so dependent on the amount of money being just right? What is the market failure that prevented the First Welfare Theorem from holding irrespective of the amount of money in the babysitting economy? ... Ponder this for a second before reading on. ... The market failure in the babysitting economy was the fact that the price of babysitting services was constitutionally fixed at one-half hour per piece of scrip. For markets to work

efficiently, prices must be free to move to equilibrate demand and supply. This was not the case in the babysitting economy. Price rigidity is the crucial “friction” that distinguishes Keynesian models of the economy—in which monetary policy plays an important role in the determination of output—from neoclassical models of the economy—in which monetary policy is much less important.

## 2 What Is Money?

In ordinary English vernacular, the word “money” is often used roughly as a synonym for “wealth.” In economics, however, “money” has a much more specific meaning. Money refers to the specific asset or object that people use to make payments when they purchase something – the medium of exchange. This is usually also the asset or object that prices and wages in the economy are posted in terms of – the unit of account.

In the U.S. economy today, virtually all prices and wages are quoted in U.S. dollars. The unit of account in the U.S. economy is therefore clearly the U.S. dollar. The situation is, however, quite a bit more complicated when it comes to the medium of exchange. U.S. dollar coins and bills are certainly used to make payment in some transaction. But this is not the case for most transactions. In most transactions, we use checks, debit cards, credit cards, and increasingly various electronic payment methods to pay for the goods and services we purchase. In these cases, we are not paying with actual U.S. dollars. Rather, in most cases, we are offering a bank’s promise to pay U.S. dollars on demand as payment. For example, when we purchase something using a debit card, funds are transferred from our checking account to the checking account of the seller. Funds in a checking account are not U.S. dollars; they are a bank’s promise to pay U.S. dollars on demand.

One traditional definition of the quantity of money in the U.S. economy is currency in circulation plus bank demand deposits held by the public. Until recently, this was the definition of the monetary aggregate called M1. But many assets other than currency and demand deposits are highly “liquid” in that they can be easily and quickly sold in exchange for currency or demand deposits. Take for instance funds you may have in a savings account or a money market mutual fund. These funds can easily and quickly be transferred into your checking account. If you need to make a large payment, one can reasonably say that the funds in your checking account, your savings account, and your money market mutual funds are available to make

that payment. For this reason, it is perhaps reasonable to include savings accounts and money market mutual funds when one seeks to measure the quantity of money in the economy. Indeed, in 2021, the Federal Reserve expanded the definition of M1 to include savings accounts (including money market deposit accounts). The Federal Reserve has traditionally also published data on a second monetary aggregate called M2. M2 is equal to M1 plus the quantity of time deposits and money market mutual funds (outside of retirement accounts). Before 2021, savings accounts were in M2 but not M1, and M2 was much larger than M1. But since 2021, the difference is minor.

There are many assets even beyond those included in M2 that are highly liquid in the sense that they trade on exchanges and can therefore be sold quickly. These assets include Treasury bills and bonds, commercial paper, stocks, and many others. One can therefore think of these assets as being available for making payments, although they may lose value and it is in some cases more time-consuming to convert them to demand deposits. In the past, the Federal Reserve published data on various monetary aggregates that were broader than M2 and therefore included some of these assets (Walter, 1989). The publication of data on these broader monetary aggregates (such as M3) has, however, been discontinued.

In addition to M1 and M2, the Federal Reserve also publishes data on a monetary aggregate called the monetary base (sometimes also referred to as high-powered money or outside money). The monetary base consists of two components: currency in circulation (i.e., notes and coins held by the public) and reserve balances (i.e., reserves that banks hold at the Federal Reserve). While currency in circulation is a component of M1 and M2, reserve balances are not. The monetary base can therefore be larger than M1 (and it sometimes has been larger).

The monetary base is an important concept because it is the monetary aggregate over which the Federal Reserve has direct control. M1 and M2 include various forms of deposits. The quantity of deposits in the economy is determined by the behavior of banks, households, and firms. People sometimes say that private banks can create money because they can issue deposits. This is true in the sense that an increase in deposits increases M1 and M2. For that reason, deposits and other components of M1 and M2 that are created by the private sector are sometimes called inside money.

Macroeconomics textbooks have traditionally described money as having a third function. In addition to being the medium of exchange and the unit of account, money can be used as a store of value. This third function of money is arguably less important than its function as a medium of exchange and a unit of account. It

will certainly play a less important role in our discussion of money and monetary economics. Most people have easy access to a multitude of other stores of value than money (i.e., the assets that make up M1 and M2). For a large fraction of households, their house is by far their largest asset by value. Many households also hold various types of financial assets – such as stocks and bonds – in retirement accounts. The wealthiest segment of society typically holds virtually all their wealth in assets other than money, for example, stocks, bonds, real estate, and shares in private businesses.

The problem with money as a store of value is that it has a very low rate of return. Currency yields a nominal rate of return of zero. Bank deposits often do pay some interest. But stocks, bonds, real estate, and many other assets yield a higher rate of return than deposits (even adjusting for risk). The reason for this is that currency and deposits provides the added benefit of being useful for transactions purposes – being liquid. Holders of money must pay for this liquidity by accepting a lower return than other assets offer. But this means that people should not hold more money than they might need for making payments. If they do, they are giving up earning a higher return on more of their wealth than they really need to.

### **3 What Role Does Money Play in the Economy?**

Money is a device to lower transactions costs. To fully appreciate this, it is instructive to start by thinking about money in a “primitive” economy (i.e., one that involves very little trade and specialization). To this end, consider a society of completely self-sufficient yeoman farmers. Each household produces everything that it consumes: they grow their own food, make their own clothes, shelter, and so on. Since there are no transactions in such a society, there is no need for money.

Now suppose that someone in this society notices that they have a comparative advantage at making shoes (say) and decides to specialize in making shoes. Rather than producing all the things that they want to consume themselves, they produce only shoes. They then exchange most of the shoes they produce for other things that they would like to consume. Over time, other people do the same. One person specializes in making bread, another specializes in teaching music, and so on. The increased specialization leads trade in this society to gradually grow. As this happens, an important issue arises: How does the trade work?

The simplest form of trade is barter: Person A has shoes but wants bread. Person B has bread but wants shoes. So, they trade shoes for bread and both are better off.



The obvious problem with this is that it may be hard for person A to find a person that has bread and wants shoes. Suppose the baker doesn't want shoes, rather he wants a music lesson. In this case, we say that there is a lack of double coincidence of wants.

One solution to the double coincidence problem is to try to arrange for a three-way barter trade. This would work in our example if the music teacher wants shoes. But the likelihood that something like this would work is low, and arranging such trades is clearly rather cumbersome in practice. A different solution is for the shoemaker to take bread as payment even if she doesn't actually want bread. Why might she do this? She might do this if she thinks that she can use the bread to buy the things that she actually wants. In other words, if she believes that others will take bread as payment, then she will be willing to take bread as payment so that she can resell it in exchange for the music lesson she desires. In this case, bread has become money – the medium of exchange in the economy.

Bread is not an ideal object to serve as money. The obvious disadvantage of using bread as money is that it is quite perishable. If bread was used as money, its perishability would imply that sellers would have a strong incentive to rush to spend the bread that they took as payment quickly before it went bad. This type of scramble to transact quickly would likely yield a great deal of inefficiency. But a clever baker might solve this problem by issuing bread tokens – i.e., coins or pieces of paper that entitled the bearer to one loaf of freshly baked bread. Such tokens could then circulate in the economy and serve as the medium of exchange.

What might be a downside of using such bread tokens as money? Actually, there are several ways in which such a monetary system might run into trouble. One problem is counterfeiting, i.e., others might create tokens that look like the tokens issued by the baker and pass them off as genuine tokens issued by the baker. A second problem is that the baker may issue too many tokens. The baker has a strong incentive to issue a large quantity of tokens since issuing tokens is likely to be highly profitable enterprise (if the baker can overcome the counterfeiting problem). But if the baker issues too many tokens, the value of the tokens may start to diminish. In addition, if the baker issues many tokens, he may not be able to honor the promise to deliver loaves of bread to those that seek to redeem his tokens. Furthermore, how many people seek to exchange their tokens for bread will depend on the confidence that the people have in the baker's ability to honor his promise. If people are highly confident in the baker's promise, they will be more likely to hold on to their tokens since they will not expect the tokens to lose value. The baker's ability to manage all

these issues will determine whether his token currency turns out to be successful.

The issues touched on in the preceding two paragraphs are some of the fundamental issues that governments, central banks, and other issuers of money (e.g., banks) have always faced in designing a monetary system. We will discuss these issues in much more detail later in this chapter and in the subsequent chapters. But before we do this, it is useful to pause and consider what characteristics are good characteristics for a medium of exchange to have. Here is a list of arguably desirable characteristics for a medium of exchange:

1. Highly standardized
2. Easily verified quality
3. High ratio of value to weight and volume
4. Easily divisible
5. Stable value

The notion that these characteristics are desirable characteristics for money to have all flow from the basic fact that the purpose of money is to reduce transactions costs. If the objects used as money are not highly standardized or their quality is not easily verified, buyers face an incentive to pay with “bad money” – i.e., low quality items. For example, if tobacco is used as money, buyers have an incentive to pay with their lowest quality tobacco. This is an example of *Gresham’s Law*: bad money drives out good money. This can be a problem even for metal coins if the weight or fineness of the coins are not standardized and are difficult to measure. Such situations require sellers to examine money carefully before taking it as payment, which increases the cost of transacting. Money that is heterogeneous or difficult to value, may also result in buyers and sellers disagreeing about the money’s value, which may result in costly negotiation. Most early forms of money suffered from these problems.

Many early forms of money were quite bulky (cows, sheep, tobacco, etc.). Clearly, it was costly for those engaged in trade to haul around these objects for the purpose of paying for the objects that they desired. Items with high value relative to their weight and volume (such as gold) have the obvious advantage that they are more convenient to carry from place to place. Paper money and ledger entries in the banking system score highly on this dimension, even relative to gold and silver.

Money should also be easily divisible. This is important so as to make it possible to carry out transactions of different size. The monetary system should ideally be

well suited for all manner of transaction from very large transactions such as wholesale trade transactions by merchants to much smaller transactions such as everyday transactions of individuals. Divisibility is particularly important for small transactions. For large transactions, one can always use many units of an object with low value (although this can be cumbersome). But if the object that is used as money has high value and is not divisible (e.g., a cow), this can make it difficult to carry out small transactions. In the U.S. today, the smallest unit of money is the penny, which is worth very little. Few things cost less than a penny. This makes it easy to carry out even small transactions. In the past, small transactions were a persistent problem as we will discuss in much more detail in section 8 below.

The final desirable characteristic of money that I list above is that money should have stable value. Using a medium of exchange with stable value contributes to stability of the prices of various goods in the economy, which arguably makes the price system more transparent and easier to use. It is important to understand in this context that the value of money is the inverse of the price level: The price level is a measure of how much money is needed to buy a certain basket of goods, which is the inverse of how many goods one can buy with one unit of money. The notion that money has stable value is therefore the same thing as the notion that the price level in the economy is stable. This implies that saying that it is desirable for money to have stable value, is the same thing as saying that inflation and deflation are costly. We will discuss the various costs of inflation and deflation in more detail in chapter XX.

When I have asked my students to list desirable characteristics of money, they sometimes suggest that it is desirable that money be in limited or fixed supply. This is an understandable suggestion. It is certainly the case that historical episodes in which money has lost large amounts of value have invariably been episodes when the quantity of money rose a great deal. In other words, large increases in the money supply can cause great instability in the value of money. But it does not follow from this that fixing the supply of money will result in money having stable value. The reason for this is that the demand for money fluctuates. If the supply of money is fixed, fluctuations in the demand for money will cause the value of money to fluctuate. An ideal monetary system is one in which the monetary authority varies the supply of money to accommodate fluctuations in the demand for money and by doing so maintains a stable value of money. This is actually a core element of modern central bank practice and one important reason why so many central banks have been so successful over the last few decades in maintaining stable inflation

rates.

But the idea that fluctuations in money demand imply that the supply of money needs to be “elastic” (i.e., needs to be able to respond to demand) if money is to have stable value is not as widely appreciated as it should be. Take for example the design of Bitcoin. Satoshi Nakamoto (pseudonym) designed bitcoin such that the supply of bitcoin is very stable. It increases gradually, but at a decreasing rate, and asymptotes to 21 million bitcoin. This design suggests that Nakamoto thought that limited supply (completely inelastic in the short run) was a desirable characteristic for a currency. One can’t argue with the success of the concept Nakamoto created. Yet the price of bitcoin is extremely unstable. This instability detracts considerably from the desirability of bitcoin as a currency. (The design of bitcoin has other problems as well (Budish, 2018).)

Let me add one final (provocative) idea to this discussion of desirable characteristics of money. This is the idea that it is desirable that money have no intrinsic value. The reason for this is simply that using objects that have other uses as money is costly. If it is possible to use intrinsically valueless objects – such as pieces of paper or electronic ledger entries – as money, this frees up the valuable objects that otherwise would be used as money for other use. This idea actually goes back all the way to a passage by Adam Smith in *The Wealth of Nations*. Smith argued that replacing gold and silver coins with bank notes would free up a large fraction of the gold and silver previously circulating for other use (not all, since Smith thought that some would need to be kept in reserve). In particular, this gold and silver could be exported in exchange for valuable imports or lent abroad in exchange for perpetual interest payments. In both cases, home consumption would rise (Smith, 1776/2000, book II, ch. 2).

## 4 A Simple Monetary Model: The Medieval Economy

A central goal of monetary economics is to understand the influence of money on output and prices in the economy. Money and monetary policy is often cited as an important cause of recessions – e.g., the Great Depression and the recession of 1982. But monetary policy has also come to be one of the primary macroeconomic policies employed to counteract recessions that arise from other causes. Furthermore, most economists believe that money and monetary policy are crucial determinants of inflation.

To understand the connection between money, output, and prices, we will over the next several chapters develop a sequence of models of money and the business cycle. In this section, we develop the first of these models. This first business cycle model will introduce a number of ideas that are central to monetary economics and Keynesian economics. However, this first model will also leave out quite a few important issues which will then be added one after another in subsequent chapters. Starting with a stripped-down model, will allow us to understand certain core ideas in as simple a setting as possible before adding additional complications.

To this end, suppose we are back in the middle ages in an economy in which the only medium of exchange is gold coins. All payments are made in gold coins. In other words, gold coins are the only object that people generally accept as payment when they transact. For simplicity, we assume that all gold coins are uniform in size, shape and fineness, or else that the value of all gold coins is proportional to the quantity of gold they contain. We will see in section 8 below that this simplifying assumption sidesteps a number of very substantial practical issues that plagued economies that used gold and silver coins as a medium of exchange through the ages.

## 4.1 Money Demand

In our medieval economy, people must hold gold coins in order to be able to purchase things. This implies that people's desire to transact results in them demanding gold coins so as to be able to transaction. In other words, there exists a transactions based demand for gold coins. (This is analogous to the demand for scrip in the babysitting economy we discussed in section 1.) People may also hold gold coins as a store of value. Finally, gold may have some (non-monetary) intrinsic value (e.g., as jewelry). These other sources of demand for gold may have played a role in gold coming to be used as money. But neither of these other sources of demand for gold play a direct role in our analysis and we therefore ignore them.

How many gold coins do people hold in the medieval economy? We assume that people's demand for gold coins – i.e., their money demand – is proportional to the nominal value of output. We represent this with the following equation

$$M_t = kP_tY_t. \quad (1)$$

In this equation,  $M_t$  denotes the number of gold coins people demand,  $Y_t$  denotes real output – i.e., output in units of goods and services as opposed to monetary

units – and  $P_t$  denotes the price level – i.e., the price of goods and services in terms of money. The product  $P_t Y_t$  is nominal output – i.e., output measured in monetary units (gold coins in this case). Finally,  $k$  is a constant of proportionality – i.e., an exogenous parameter.

In a monetary model such as the one we are developing, it is important to understand the distinction between nominal variables and real variables. Nominal variables are measured in monetary units—such as dollar, pounds sterling, yen, or gold coins. Real variables, on the other hand, are measured in physical units. Consider for example an economy in which the only good produced is potatoes. The real value of output in such an economy is measured in kilograms of potatoes (or some other weight unit). The nominal value of output, however, is measured in dollars (or some other monetary unit) and is equal to the real value of output (kilograms of potatoes) times the price of potatoes (dollars per kilogram of potatoes). In real world economies, the monetary unit typically has a name (e.g., dollar). In our medieval economy, however, the monetary unit is simply gold coins.

Why do we assume that money demand is proportional to nominal output? This assumption is motivated by the idea that money demand is increasing in the value of transactions people engage in and the value of transactions people engage in is increasing in the value of output in the economy. Intuitively, since people hold money to be able to transact, it seems reasonable to assume that people hold more money the larger is the value of the transactions they need to engage in. It also seems reasonable to assume that the value of transactions people engage in is larger the larger is the amount of output produced in the economy.

Consider a very simple economy in which people are paid once a month and use that money to pay for the things that they purchase until their next pay day. If their expenditures are evenly spaced over the month, people in this economy will end up holding  $1/2$  a month's worth of their wages (and purchases) in money on average over the month. This super-simple example abstracts from many features of reality. For example, people may decide to deposit some of their earnings in the bank when they get paid and then withdraw these funds later in the month. This would reduce their average money holdings over the month. (Here I am excluding bank deposits from my definition of money in this simple economy. Alternatively one could imagine they purchasing some other asset.) On the other hand, they may want to hold some extra money in case of an emergency of some sort. This would increase their money holdings. Whether the amount of money people decide to hold is equal to  $1/2$  a month of output, more, or less, will thus depend on factors such as

the ease with which they can buy and sell assets that are superior stores of value to money and the uncertainty they face regarding the quantity of transaction they will engage in. In the language of our model, these factors determine the parameter  $k$ .

Equation (1) is sometimes written in a slightly different way as

$$M_t V = P_t Y_t. \quad (2)$$

This is the same equation as equation (1). The only difference is that we have defined  $V = 1/k$  and multiplied both sides of equation (1) by  $V$ . In other words, equation (2) is simply equation (1) with the exogenous parameter written in a different form. Equation (2) is usually referred to as the *quantity equation* and is an integral part of the so-called quantity theory of money. The medieval model we develop in this chapter is thus a quantity theoretic model of money and money's influence on output and prices. The logic of these names will hopefully become clear later in the chapter. They derive from the fact that changes in the quantity of money  $M_t$  play a central role in this model.

The exogenous parameter  $V$  is usually referred to as the “velocity” of money. The idea behind this name is that  $V$  measures the number of times each gold coin needs to change hands over some period of time. For example, if  $M_t = 10$  but  $P_t Y_t = 100$  (over a time period such as a year) then the idea is that each gold coin needs to change hands 10 times (over that period). While somewhat intuitively appealing, this notion, if taken too literally, risks encouraging an overly simplistic view of the quantity of transactions in an actual economy.

In reality, the quantity of transactions is vastly larger than the value of output. One reason for this is that the production of many products is divided into several stages with separate firms specializing in each stage of production. This implies that there is a large amount of trade in intermediate inputs in actual economies (and consequently a large volume of associated transactions). Another large category of transactions is factor payments such as payments of wages. In the U.S. economy in 2011, the volume of non-financial transactions equaled \$71 trillion, while GDP was only \$15 trillion. But the volume of non-financial transactions turns out to be tiny compared to the truly gargantuan volume of financial transactions in modern economies (i.e., assets being bought and sold). In 2011, the volume of financial transactions in the U.S. exceeded \$2,000 trillion (Piazzesi and Schneider, 2018).

Let's assume that our medieval economy consists of a large number of identical households. For simplicity, assume that there is actually a continuum of such households in the economy of length one. The population in our economy is then equal to

one. But this does not mean that there is a single household in the economy. Rather it means that there is an infinite number of households of “length” or “size” one. In this case, equations (1) and (2) describe not only money demand for a single household but also aggregate money demand in the economy. I.e.,  $M_t$  is both per capita money demand and total money demand. This is the case because total money demand is equal to per capita money demand multiplied by the size of the population, which is one. The same is true of  $Y_t$ . It is both per capita output and total output in the medieval economy.

What about the supply of gold coins in our medieval economy? We assume that the supply of gold coins is given exogenously. In other words,  $M_t$  is an exogenous variable in the medieval economy. For now, our baseline assumption is that the quantity of gold coins in circulation does not change except on rare occasions. Think of our medieval economy as being an island economy that does not trade with the rest of the world and has no gold mines. Most of the time, the quantity of gold coins remains unchanged.

To understand the role that equations (1) and (2) play in our model of the medieval economy, it is useful to view them from two different perspectives. From the perspective of a single household, these equations describe money demand. I.e., they describe how much money a single household chooses to hold as a function of the nominal value of per capita output in the economy. However, from the perspective of the entire economy, the supply of money is given. In aggregate, the households in the economy, therefore, cannot choose how much money to hold freely. They must end up collectively holding the amount of money that actually is circulating in the economy. This means that from the perspective of the entire economy, the most useful way to think about equations (1) and (2) is as equations that help determine the values of  $P_t$  and  $Y_t$  given the exogenous value of  $M_t$ . Later in the chapter, we will discuss in greater detail how these equations can be thought of as (two versions of) the aggregate demand curve in our medieval model.

## 4.2 Production and Price Setting

In the medieval economy each household owns a small business such as a bakery, a pub, a farm, or a shoe shop. For simplicity, we assume that households produce output with only labor and the production function is linear in labor:

$$Y_t = AL_t. \tag{3}$$



Here  $L_t$  denotes labor and  $A$  denotes productivity. We assume that productivity is exogenous.

The way trade works in the medieval economy is that each business posts a price for its product at the beginning of each period (before they observe demand that period). The business then stands ready to service all customers that demand their good at this price over the course of the period. For now, think of the period as a day. The business posts a price before opening each day and stands ready to service all customers that want to purchase its product on that day at that price.

These two assumptions – posted prices that are unresponsive to demand in the short run and a commitment from sellers to meet demand at these posted prices – are the core assumptions of Keynesian models. It is these assumptions that make the medieval economy model a Keynesian model. In a competitive markets model, prices are not posted in advance; they are determined by the intersection of demand and supply. In a Keynesian model, however, each producer is a monopolist supplier of their unique good and therefore have the power to choose the price of their good. Once the producers have set these prices and committed to supply however much of their goods is demanded at these prices, output is determined purely by demand. If demand is high at the posted prices, output will be high. If demand is low at the posted prices, output will be low. This means that in a Keynesian model output is purely demand determined in the short run.

If demand ends up being high on a particular day, this means that a large amount needs to be produced over the course of that day, which in turn means that the producers need to work hard on that day. Consider, for example, a shoe maker. If few customers show up, the shoe maker does not need to work very hard since she doesn't need to make many shoes. If however many customers show up, the shoe maker needs to make many shoes which means she needs to work hard.

Working hard has both costs and benefits. Hard work results in higher income. But it is also exhausting. Furthermore, as we discussed in more detail in chapter XX [Labor Chapter], the marginal utility of income falls as income rises, while the marginal disutility of work effort rises as work effort rises. This implies that more work is not always a good thing. There is some level of work effort beyond which people's overall utility starts to fall because the marginal utility of the income they earn is not large enough to compensate for the marginal disutility of labor. In other words, there exists an ideal or desired level of labor. We denote this desired level of labor  $L^*$ . People do not want to work more than  $L^*$  because at that point, the extra effort is sufficiently costly to them that the extra income they earn from their extra

effort is not worth the trouble.

Since producers post prices in advance and commit to meet demand, they may end up having to work more than they desire. Think of the shoe maker again. She posts her prices at the beginning of the day. Then as the day wears on she may find that many more customers frequent her business and buy shoes than she had expected. By the end of the day, she may be completely exhausted to the point where she thinks to herself: “I need to do something to put a lid on demand for my shoes.”

What can she do? One thing she can do is to raise her prices. This will reduce demand for her shoes and thereby reduce her labor effort. But how much should she raise her prices? One thing she may worry about as she ponders this question is whether the high demand she has experienced was something temporary – perhaps due to a swing in fashion towards her particular type of shoe that may not last. If she raises her prices and the high demand she experienced turns out to have been temporary, she may end up with too low demand the next day.

These types of concerns may motivate her to adopt a cautious approach where she raises her price a little bit when she experiences high demand. If demand continues to be high the next day, she then raises her price a little bit more. She does this until demand for her shoes has fallen enough that her labor supply has reached its desired level of  $L^*$ .

We assume that all producers in our medieval economy adopt this approach to managing demand for their products. We formalize this behavior with the following price setting equation:

$$\frac{P_{t+1}}{P_t} = \left( \frac{L_t}{L^*} \right)^\theta . \quad (4)$$

This equation implies that if labor supply is above its desired level  $L_t > L^*$ , then prices will rise, i.e., producers will set higher prices next period than this period ( $P_{t+1} > P_t$ ). Likewise, if labor is below its desired level, prices will fall.

The parameter  $\theta$  (Greek letter “theta”) determines the speed of price adjustment. If  $\theta$  is large, prices respond strongly to deviations of labor supply from its desired level. If  $\theta$  is small, the response of prices is more sluggish. We described one reason why prices may respond sluggishly to demand above. This was the idea that producers are unsure whether the high demand they face is temporary or permanent and therefore adopt a gradual approach to price adjustment. But this idea is not likely to be the whole story when it comes to sluggish price adjustment in the real world. There seems to be something more at play.

It is a familiar fact about prices that they do not change very often (unless you live in a country suffering from high inflation). For example, think about how long it has been since your hair dresser changed their price. If you frequent a restaurant, consider how long it has been since the prices of the dishes you most often order have changed. Many prices at the grocery store and other establishment also remain unchanged for considerable periods of time. At first glance, this may seem natural. But is it? The demand these producers face is constantly changing, as are the costs of their inputs. In competitive markets, if demand and costs are constantly shifting around, prices will be in constant motion. In reality, however, many prices change very infrequently. They seem to be sticky or rigid.

This stickiness of prices – i.e., the fact that prices often don't change *at all* for long periods of time – is another reason why prices respond sluggishly to changes in demand and costs. In other words, price stickiness (a.k.a. price rigidity) is an important determinant of the value of the parameter  $\theta$  in our price setting equation – equation (4). For this reason, macroeconomists have devoted a great deal of effort to measuring price stickiness. Table 1 present estimates of price stickiness from a paper of mine co-authored with Emi Nakamura. The table reports the median duration of consumer prices for various categories of products in the United States between 1998 and 2005. The first thing to note about the results in this table is that the degree of price rigidity varies a great deal across product categories. On the one hand, prices of services are very sticky. The median duration in the services category is 12.9 months – i.e., prices change less than once a year. On the other hand, prices of vehicle fuel, travel, utilities, and transportation goods (e.g., cars) are quite flexible. They change almost every month (or more than once a month in the case of vehicle fuel). When we consider all products, the median duration of prices is roughly 9 months.

But wait. Table 1 actually reports two numbers for each category. The first column reports the median duration of “regular prices,” while the second column reports the median duration of raw prices. For several product categories and for the median over all products, this distinction is quantitatively quite important with raw prices being much more flexible than “regular” prices. Which numbers should we be paying attention to?

Figure 1 illustrates what the distinction between regular prices and raw prices is about. This figure plots the price of Nabisco Premium Saltines – a typical super-market product – over a roughly seven year period. A dramatic feature of this price series is that the price of Saltines often drops by a large amount for a brief period

Table 1: Price Rigidity in the United States

	Median Duration of Prices	
	Regular Prices	Raw Prices
Services	12.9	12.8
Household Furnishing	10.2	3.8
Recreation Goods	10.0	5.5
Apparel	8.8	2.2
Processed Food	8.5	3.2
Other Goods	5.9	5.3
Unprocessed Food	3.4	2.1
Transportation Goods	2.2	2.2
Utilities	2.1	2.1
Travel	1.8	1.7
Vehicle Fuel	0.5	0.5
All	8.7	4.4

*Notes:* These estimates are calculated using data from Tables I and V in Nakamura and Steinsson (2008). The median duration of prices is calculated as  $d = -1/\ln(1 - f)$  where  $f$  is the median frequency of price change in months. The estimates of the frequency of price change include substitutions. The service category excludes travel. The sample period is 1998-2005.

and then reverts right back to exactly the price it was at before. These short-lived price reductions are what we usually refer to as the product “going on sale.” Such temporary sales are quit frequent for many products. As a result, the prices of these products change quite often. For example, the price of Saltines changed 117 times over the 365 week period plotted in Figure 1. Given this, perhaps we should view Saltines as having a very flexible price.

Not necessarily. The price changes associated with temporary sales are very peculiar in that they have no long-lasting effect on the price of the product. When the sale ends, the price of the product is right back where it started. This contrasts sharply with what are usually referred to as “regular price changes.” Let’s define the price of Saltines when it is not on sale as the “regular price.” In Figure 1, we see that the regular price of Saltines changes rather infrequently. There are only nine changes in the regular price over this seven year period. If we focus on these, we might say that the price of Saltines (that is the regular price) is quite sticky.

How rigid one views prices to be, thus, depends critically on how one handles temporary sales. What to do in this regard is somewhat controversial. However, in

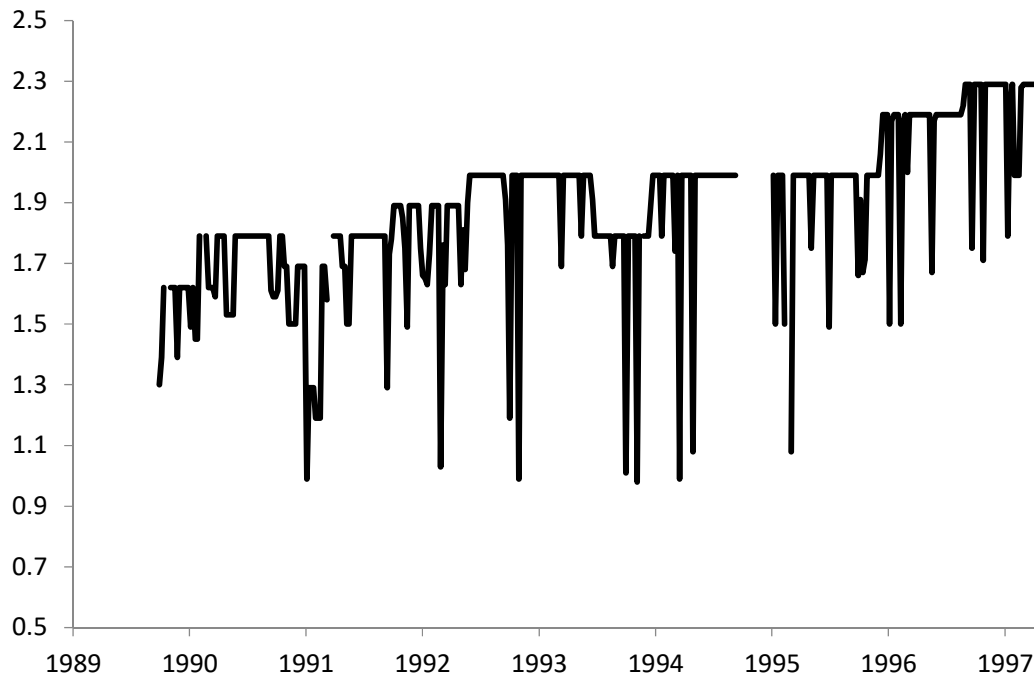


Figure 1: Price of Nabisco Premium Saltines (16 oz)

*Note:* Price at Dominick's Finer Food store in Chicago.

my view, a strong case has been made in the literature on price rigidity that temporary sales do not contribute much to the response of prices to changes in demand and costs. This is the case for two reasons. First, prices revert quickly to their prior value, implying that temporary sales have no long-lasting effect on the level of prices. The second reason is that most temporary sales occur for reasons that are unrelated to demand and costs. Arguably, most temporary sales occur for reasons of marketing and price discrimination. Nakamura and Steinsson (2013) discuss these ideas in more detail.

Why do firms choose not change their prices more often? Economist Alan Blinder and co-authors attempted to answer this question by simply asking firm managers (Blinder et al., 1998). When they asked the open-ended question: Why don't you change prices more frequently? the most common answer given by firm managers was that they were worried about antagonizing their customers. Other frequent answers were that they worried about competitive pressure, that changing prices is costly, and that their costs don't change more often.

Blinder and co-authors also described to the managers in plain English the logic of twelve different theories economists have proposed for price stickiness and asked them how important these ideas were for their firm. The idea that resonated most

strongly with firm managers was that they would like to change their price but were reluctant to do this because changing their price would lead their price to get out of line with the prices of their competitors. This is an example of a *coordination failure*: everyone wants to change their price, but only if everyone else does the same. Since everyone is uncertain what others will do, no one changes their price.

Another idea that resonated strongly with the firm managers Blinder and co-authors surveyed was the idea of *implicit contracts* between firms and their customers to maintain stable prices. According to this idea, firms have an implicit understanding with their customers that they will not take advantage of periods of strong demand by raising their prices. One way to understand the logic for such implicit contracts is as an insurance arrangement: customers value being insured against price increases when demand is high. Another possible reason for implicit contracts is that they represent a commitment by firms not to take advantage of customers who have become partially locked into purchasing their product (e.g., because switching to a competing product is costly).

A third idea that scored highly with the firm managers Blinder and co-authors surveyed was the idea that it was better to adjust other product characteristics than the price. For example, firms might respond to an increase in demand by reducing advertising and other sales efforts, increasing delivery lags, reducing service, or reducing the quality of the product. When asked why they preferred responding in these ways, the most common answer was that they believed that these responses were less costly and less likely to antagonize their customers.

## 5 Vikings Bring Back Gold Plunder

Suppose our medieval economy sends off a ship of vikings to plunder gold from unsuspecting monasteries in neighboring countries. After their raiding is done, the vikings come home with their ship loaded to the brim with gold coins. The questions we are interested in is how the arrival of this gold affects the medieval economy. More specifically, we would like to understand what happens to output and prices in the medieval economy in the short run, as time passes, and in the long run.

Let's start by considering what might happen without the aid of any equations. Initially, the vikings and whoever financed their expedition are likely to go on a spending spree. They will spend some of their gold plunder on consumer goods (beer, mutton, new shoes, a new battle ax, etc.). But they may also use some of it

to purchase assets (a bigger farm, more cattle, a bigger house, etc.). This spending leads the gold they brought back to start diffusing through the economy. As this happens, others in the economy find that they have more gold coins, which leads them to increase their spending as well. Before long, everyone is spending more than before. In other words the medieval economy experiences an economic boom.

Notice that each time one person spends a gold coin another person ends up holding an extra gold coin. The people in the medieval economy are spending more than before because they have more gold coins than before. But their spending is not directly dissipating this situation. The extra spending is just moving the gold around the economy. Economists often refer to this situation as one in which too much money is chasing too few goods.

The economic boom does, however, imply that the producers in the economy end up working more. Since the economy was in steady state before the vikings returned, the extra work to meet the extra demand means that the producers in the economy are working more than they would like. They respond to this by starting to raise their prices. As prices rise, the purchasing power of the gold coins people hold – their “real value” – falls, which is likely to reduce people’s desire to spend.

To summarize, this discussion conjectures that the short run response of the economy to an increase in the money supply is an economic boom. The economic boom then leads producers to start raising their prices. As prices rise, the economic boom is diminished. But where does the economy end up in the long run? The logic outlined above suggests that prices will rise until the real value of the money supply has fallen back to its original value – i.e., until prices have risen by the same proportion as the money supply increased when the vikings brought back the extra gold. At that point, people are holding the same amount of money in real terms as before and therefore demand and output will be back to normal and prices will stop rising.

Let’s now see whether this description of events is what actually happens according to our medieval economy model. The first step towards this is for us to simplify the model. Above, we defined the desired level of labor supply as  $L^*$ . We now analogously define the desired level of output as the output produced with the desired level of labor:  $Y^* = AL^*$ . We can then use this definition of  $Y^*$  and the production function to rewrite the price setting equation – equation (4) – in terms of deviations of output from its desired level as opposed to deviations of labor supply from its desired level:

$$\frac{P_{t+1}}{P_t} = \left( \frac{Y_t}{Y^*} \right)^\theta . \quad (5)$$

With this simplification, the medieval economy model consists of only two equations: 1) the price setting equation (equation (5)), and 2) the quantity equation (equation (2)). (Recall that equations (1) and (2) are just two ways of writing the same equation.) The model now has two endogenous variables:  $Y_t$  and  $P_t$ . The money supply  $M_t$  is an exogenous variable, and the remaining parameters and variables ( $V$ ,  $Y^*$ , and  $\theta$ ) are also exogenously given.

## 5.1 Long-Run Monetary Neutrality

Let's suppose for simplicity that no changes in the quantity of gold coins had occurred for a long while before the arrival of the viking ship, and the economy had therefore settled down to a steady state in which both prices and output were constant. We denote the value of each variable in this initial steady state simply by removing its time subscript. So, steady-state output and prices before the arrival of the viking ship are denoted  $Y$  and  $P$ , respectively.

In this initial steady state – since prices are not changing over time – the left-hand-side of equation (5) becomes  $P/P = 1$ , which implies that the equation becomes  $1 = (Y/Y^*)^\theta$ . Manipulation of this equation yields  $Y = Y^*$ . In other words, output is at its desired level in the initial steady state.

Given this results, we can use the quantity equation – equation (2) – to solve for the steady state price level prior to the arrival of the viking ship. If we denote the quantity of gold in the economy prior to the arrival of the viking ship as  $M$ , then the steady state of the quantity equation is  $MV = PY$ . We can then plug in  $Y^*$  for  $Y$  and divide through by  $Y^*$  to get that  $P = MV/Y^*$ . Notice that we have now solved for the steady state of both output and prices in terms of only exogenous variables and parameters.

When the viking ship arrives, the money supply in our medieval economy increases abruptly from  $M$  to a larger value which we denote by  $\tilde{M}$  (M-tilde). If we wait long enough after the viking ship arrives, the economy will settle down to a new steady state. We denote the value of output and prices in this new steady state by  $\tilde{Y}$  and  $\tilde{P}$ , respectively. The same argument as above implies that in the new steady state  $\tilde{Y} = Y^*$  and  $\tilde{P} = \tilde{M}V/Y^*$ .

In the long run, therefore, the extra gold has no effect on output ( $\tilde{Y} = Y = Y^*$ ). The only effect that the extra gold has on the economy in the long run is to raise prices. How much exactly do prices rise? We can see this by dividing the expression



for  $\tilde{P}$  by the expression for  $P$ . This yields:

$$\frac{\tilde{P}}{P} = \frac{\tilde{M}V/Y^*}{MV/Y^*} = \frac{\tilde{M}}{M}. \quad (6)$$

This equation says that in the long run the price level increases by the same proportion as the money supply. If the amount of gold that the vikings brought back raised the money supply by 20%, equation (6) shows that the price level will eventually also rise by 20%. When the price level has risen by the same proportion as the money supply, the real value of the money supply is back to its original value. We can see this by manipulating equation (6) to yield  $\tilde{M}/\tilde{P} = M/P$ . It is at this point that demand for goods in the economy returns to its original value.

The results derived above are some of the most basic results in monetary economics. They are often referred to as the *Classical Dichotomy* or as *Long-Run Monetary Neutrality*. The Classical Dichotomy says that *in the long run* changes in the money supply have no effect on “real” variables (such as output); they only affect “nominal” variables (such as the price level). The Classical Dichotomy holds in our medieval economy model in response to a one-time increase in the money supply. When the Classical Dichotomy holds, we say that changes in the supply of money are “neutral” in the long run (i.e., don’t affect real variables). We will see in chapter XX [Phillips Curve chapter] that there are ways to break the Classical Dichotomy in our simple medieval economy model.

The idea for why economists think the Classical Dichotomy should hold (or at least should be close to holding) is that choosing the quantity of money in the economy is somewhat similar to choosing a unit of measure. Just as it should not matter for anything whether we measure weight in kilograms or pounds, it should not matter whether we measure prices in dollars or in yen (which are worth roughly 100 times less than a dollar as of this writing). If the money supply increases by a factor of 100, all prices should simply increase by a factor of 100 (in the long run) and nothing “real” should be affected.

## 5.2 Short Run Monetary Non-Neutrality

We now turn to the short-run effects of the arrival of the viking gold on the medieval economy. To analyze these effects, it is useful to take logs of the equations of the model. Taking logs of the quantity equation – equation (2) – yields

$$\log M_t + \log V = \log P_t + \log Y_t. \quad (7)$$

(Recall that in this book  $\log$  refers to the natural logarithm.) Taking logs of the price setting equation – equation (5) – yields

$$\log P_{t+1} - \log P_t = \theta(\log Y_t - \log Y^*). \quad (8)$$

The medieval economy model is a dynamic model (just as the Malthus model analyzed in chapter XX and the Solow model analyzed in chapter XX are dynamic models). The price setting equation – equation (8) – is the dynamic equation in the model, i.e., the equation that links events in different time periods. Despite being a dynamic model, the medieval economy model is relatively simple to solve. The solution involves only iterating back and forth between the price setting equation and the quantity equation.

In working through these dynamics, it will prove convenient to use a version of the quantity equation written in terms of changes as opposed to levels. To this end, we subtract the quantity equation for time  $t - 1$  from the quantity equation for time  $t$ . This yields

$$\log M_t - \log M_{t-1} + \log V - \log V = \log P_t - \log P_{t-1} + \log Y_t - \log Y_{t-1}$$

which we can write as

$$\Delta \log M_t = \Delta \log P_t + \Delta \log Y_t. \quad (9)$$

Here we are using the symbol  $\Delta$  (Greek letter capital “delta”) to denote a “first difference”, i.e., a change from one period to then next. So, for a variable  $\log X_t$  we have that  $\Delta \log X_t = \log X_t - \log X_{t-1}$ . Since we are assuming that velocity is constant, it drops out when we write the quantity equation in changes.

Let’s refer to the period in which the gold arrives as period 0. Since we are assuming that the economy was in steady state before the gold arrives, we know that  $\log Y_{-1} = \log Y^*$ , i.e., output in period -1 (the period before the gold arrives) was equal to its desired level. We can furthermore use the quantity equation to solve for  $\log P_{-1} = \log M + \log V - \log Y^*$ , where we are using the fact that  $\log Y_{-1} = \log Y^*$ .

Next, consider how the economy reacts at time 0, the date on which the gold arrives. Since prices in the medieval economy are set one period in advance, they will be unchanged: At the end of period -1, price setters take stock of what happened that period and decide on prices for period 0. Since nothing out of the ordinary happened in period -1, and output turned out to be equal to its desired level, price setters decide not to change their prices between period -1 and period 0. We can

formally derive this by observing that the price level in period 0 is determined using the price setting equation – equation (8) – with  $t = -1$ :

$$\log P_0 - \log P_{-1} = \theta(\log Y_{-1} - \log Y^*) = 0,$$

which implies that  $\log P_0 = \log P_{-1}$ .

With  $\log P_0$  in hand, we can use equation (9) to solve for  $\log Y_0$ . Equation (9) (with  $t = 0$ ) implies that

$$\Delta \log Y_0 = \Delta \log M_0 - \Delta \log P_0.$$

We have seen above that  $\Delta \log P_0 = \log P_0 - \log P_{-1} = 0$ . So, we have that  $\Delta \log Y_0 = \Delta \log M_0$ . This implies that the proportional increase in output in period 0 is equal to the proportional increase in the money supply. (Recall that  $\Delta \log Y_0 = \Delta \log M_0$  implies  $Y_0/Y_{-1} = M_0/M_{-1}$ .) Furthermore, to a first order approximation, the percentage change in output is equal to the percentage change in the money supply. Going forward, we will often refer to log changes and percentage changes interchangeably when the changes in question are not very large.

We thus see that our medieval economy has the simple implication that, on impact, prices do not change and output rises by the same proportion as the money supply. In this sense, “the entire” increase in the money supply shows up in output in the very short run.

Let’s now move on to period 1. The tactic we use to solve the model is the same for period 1 and for period 0. First, we use the price setting equation (with  $t = 0$ ) to solve for  $\log P_1$  given  $\log Y_0$ . Then we use the quantity equation (with  $t = 1$ ) to solve for  $\log Y_1$  given  $\log P_1$ . At the end of period 0, price setters again take stock of their situation. Since demand was higher than its desired level in period 0, they decide to raise prices. Formally, they set

$$\log P_1 - \log P_0 = \theta(\log Y_0 - \log Y^*).$$

Since  $\log Y_0 - \log Y^* = \Delta \log Y_0 = \Delta \log M_0$ , we have that  $\Delta \log P_1 = \theta \Delta \log M_0$ .

Armed with  $\Delta \log P_1$ , we can use equation (9) (with  $t = 1$ ) to get that

$$\Delta \log Y_1 = \Delta \log M_1 - \Delta \log P_1 = -\theta \Delta \log M_0.$$

Here we make use of the fact that there is no change in money supply between period 0 and period 1:  $\Delta \log M_1 = 0$ . The level of output relative to its desired level

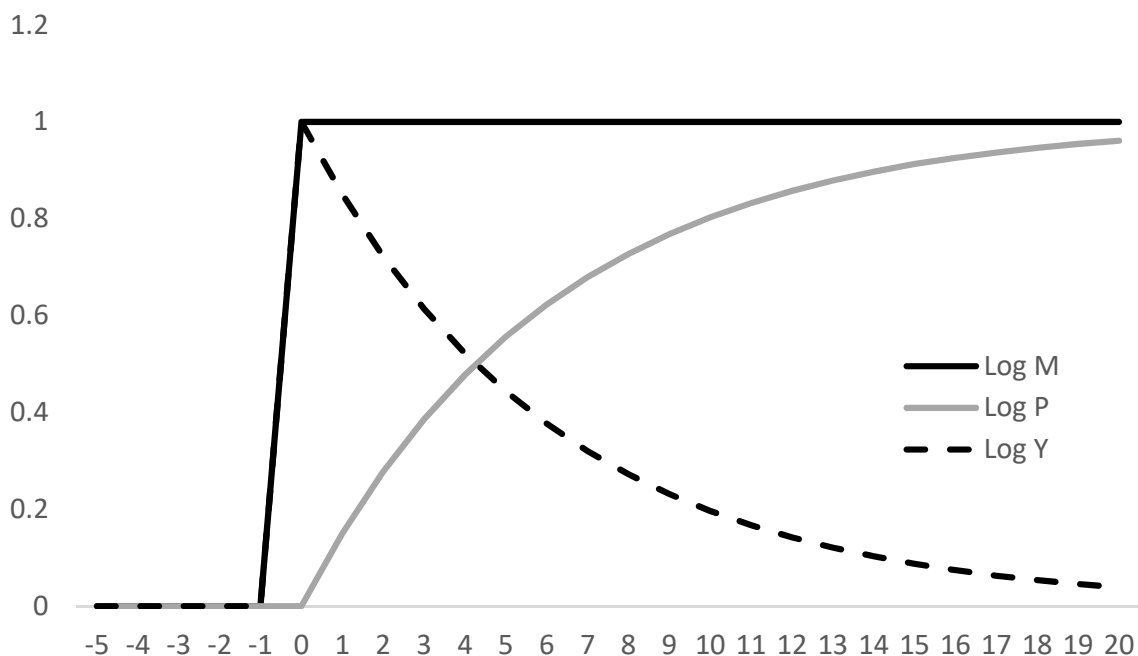


Figure 2: Response of Medieval Economy to an Increase in the Money Supply

is

$$\begin{aligned}
 \log Y_1 - \log Y^* &= \Delta \log Y_1 + \Delta \log Y_0 \\
 &= -\theta \Delta \log M_0 + \Delta \log M_0 \\
 &= (1 - \theta) \Delta \log M_0.
 \end{aligned}$$

Output falls in period 1, but remains above its desired level.

These derivations show that in the period after the gold arrives ( $t=1$ ), prices begin to adjust and output begins to return to its steady state. The speed of this adjustment process is governed by the parameter  $\theta$ , i.e., by how sticky prices are in the medieval economy. The stickier are prices – i.e., the smaller is  $\theta$  – the slower is the adjustment of the economy to its new steady state.

It is straightforward to continue solving for output and prices in periods 2 and beyond. As in periods 0 and 1, this can be done by using the price setting equation to solve for prices given output in the previous period and then by using the quantity equation to solve for output given prices. Rather than work through these steps for more periods, we present the entire path of output and prices as a function of time after the arrival of the viking ship in Figure 2.

In constructing Figure 2, we make specific assumptions about the exogenous variables and parameters of the model. To make the figure as simple as possible,

we assume that  $\log Y^* = 0$ ,  $\log V = 0$ , and  $\log M = 0$ . This implies that both the logarithm of output and the price level are equal to zero before the arrival of the viking ship. We then assume that the arrival of the viking ship raises the logarithm of the money supply from 0 to 1, i.e.,  $\log \tilde{M} = 1$ . Finally, we set  $\theta = 0.15$ .

As we had derived analytically, the initial response of the economy is for the logarithm of output to rise by the same amount as the logarithm of the money supply (i.e.,  $\log Y_0 = 1$ ) and for the price level to remain unchanged. In subsequent periods, prices gradually rise and output gradually falls. As time passes, the logarithm of the price level asymptotes to one, i.e.,  $\log P_t \rightarrow 1$  as  $t \rightarrow \infty$ , and the logarithm of output asymptotes to zero, i.e.,  $\log Y_t \rightarrow 0$  as  $t \rightarrow \infty$ . The speed of these dynamics is governed by  $\theta$ . Each period, prices rise  $\theta$  fraction of the way they still have to go to get to one and output falls  $\theta$  fraction of the way it still has to go to get to zero.

We have now seen that the arrival of the viking plunder leads to a short term boom in output. Output rises because prices are slow to adjust to the change in the money supply. As time passes, prices gradually rise and output falls back to normal. In the long run, output is unaffected by the viking plunder and prices are permanently higher.

Did the arrival of the viking plunder make the people living in the medieval economy better off? Many people's first reaction is to think that it did in the short run since output rose. This is however not correct. It is important to remember that the workers in the economy needed to produce the extra output. Whether people are better off depends on whether the extra output yields more utility than the extra effort needed to produce that output yields disutility. In the medieval economy, we have assumed that output is at its desired level in steady state. What the monetary injection does, therefore, is to raise output above its desired level. This actually lowers people's welfare (they wish output were lower since this would mean less work).

Of course, the effect of the viking plunder affects different people in the medieval economy differently. The vikings themselves became wealthier by plundering the gold. So, they were better off. But everyone else just worked more than they want and were made worse off. This example illustrates well how striking gold may be good for the individual, but not necessarily good for society.

Before moving on to the next topic, let me add a little bit of nuance to the conclusions we came to in this section. In particular, the conclusion that a boom in output that results from a monetary injection reduces welfare depends critically on the assumption that we made above that the economy was operating at its efficient level

in steady state (we called it the desired level of output). There are, however, reasons to believe that steady state output in real world economies may be below its efficient or desired level. If the producers in the economy have market power, they will likely find it in their interest to set the prices of the goods they produce above marginal cost. This will imply that output in the economy will end up being lower than is efficient. In such an economy, a moderate boom in output that results from a monetary injection will raise welfare because output is temporarily brought closer to its efficient level by the boom.

## 6 The Price Revolution

During much of the last millennium, the unit of account throughout most of Europe was defined in terms of silver and gold coins. For example, in England a silver penny coin was minted for over a thousand years as well as various other silver and gold coins with values defined in terms of pennies, shillings, or pounds. England was therefore on a specie standard – a bimetallic standard early on and later a gold standard. Today, few currencies are on a specie standard of any kind. Some countries peg their currencies to other currencies – typically the U.S. Dollar or the Euro. However, the major currencies of the world – U.S. Dollar, Euro, Japanese Yen, Chinese Renminbi, British Pound, etc. – are not backed by anything: they are pure fiat currencies.

An important potential worry with operating a fiat currency is that the value of the currency may not remain stable. Throughout the fiat currency era, some economic commentators as well as politicians (typically on the right of the political spectrum) advocate the return to the gold standard on the grounds that this will prevent inflation and result in a more stable value of the currency. Over the next few chapters, we will discuss this policy proposal several times and make use of a wide variety of evidence and theoretical ideas to assess it. We start this discussion here by considering the evolution of the price level in England over a 500 year period from the late middle ages until right before the Industrial Revolution.

Figure 3 plots the evolution of the price level in England from 1260 to 1750. The price level is plotted in two different units. The gray line plots the price level in pounds sterling – the unit of account in England. The black line plots the price level in grams of silver. Let's begin by focusing on the gray line. A striking feature of the gray line is how stable it is from 1260 until 1500. For about 250 years, there

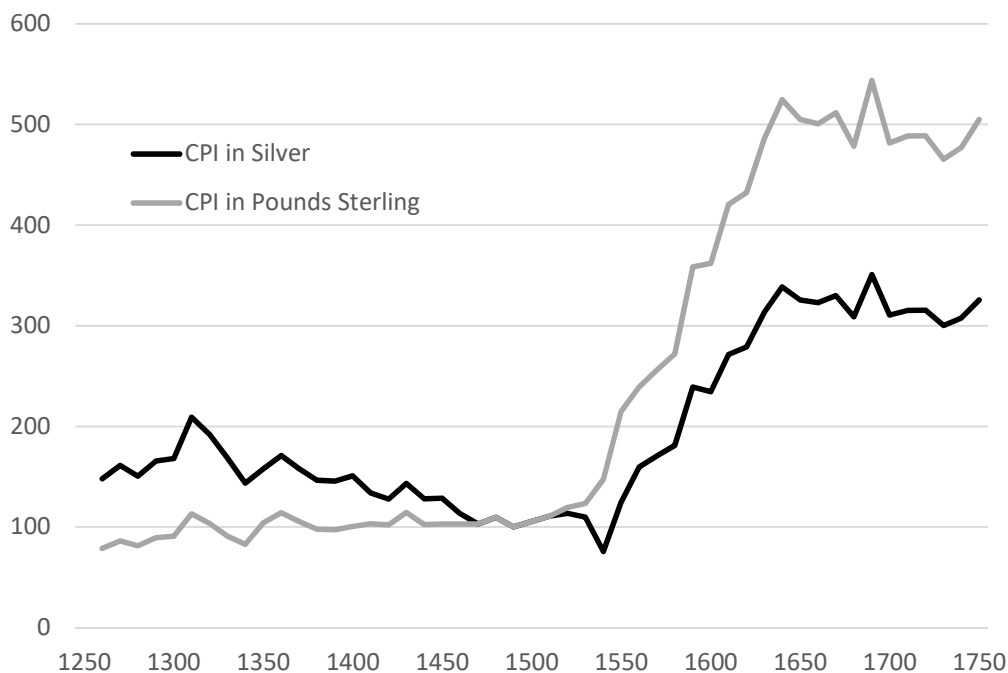


Figure 3: Price Level in England from 1260-1750

*Note:* The figure plots an estimate of the consumer price index for England from 1260 to 1750. The black line is consumer prices denominated in grams of silver. The gray line is consumer prices denominated in pence (i.e., in pounds sterling). These estimates are from Allen (2001).

was virtual price stability in England. But then something happened and prices started to rise. Over the next 150 years, prices in England rose by a factor of five before stabilizing and remaining relatively stable from 1650 to 1750. The evolution of prices in other parts of Europe was similar. All countries in Europe for which data exist experienced a large increase in prices from 1500 to 1650. The size and scope of this event has led in economic historians to refer to it as the *Price Revolution*.

Clearly, being on a specie standard does not guarantee price stability: the value of the pound sterling fell by 80% over the 16th and first half of the 17th centuries. What caused this huge increase in prices? The theoretical discussion earlier in this chapter suggests a possible explanation: Perhaps Europe experienced a major increase in the money supply that led prices to rise. As it turns out, Europe did in fact experience a massive inflow of gold and silver (mostly silver) starting around 1500. Following Columbus' discovery of the New World in 1492, the Spanish began importing treasure from the Americas. Initially, they found only modest quantities of gold and silver. But following the discovery of rich silver deposits at Guanajuato in Mexico and Potosi in modern day Bolivia, these imports went from being a trickle

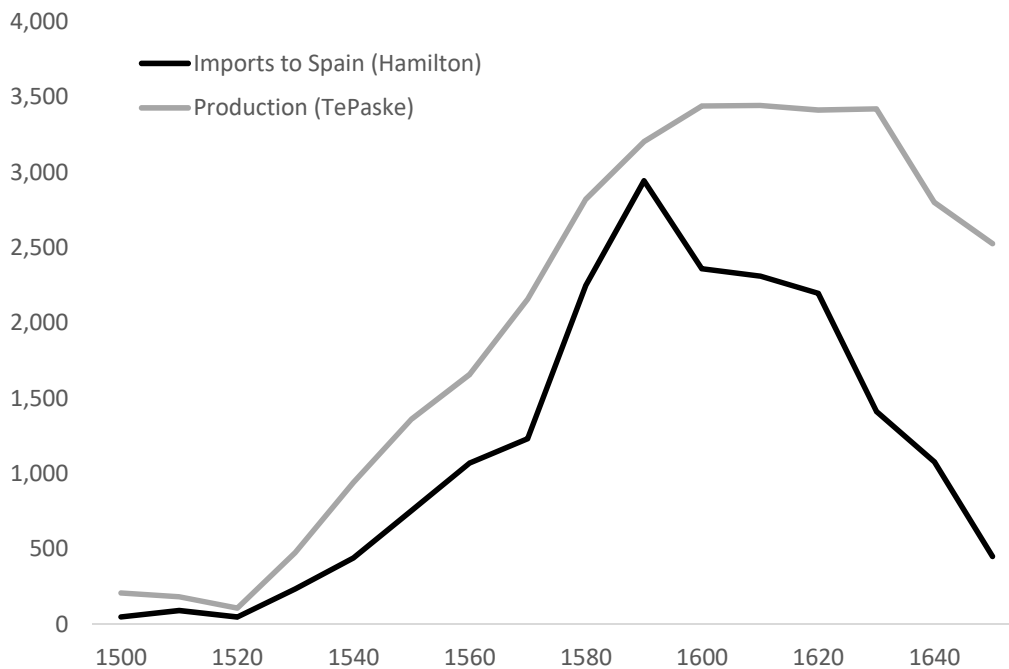


Figure 4: American Treasure: Production and Imports to Spain

*Note:* The black line plots Hamilton's (1934) estimates of imports of gold and silver to Spain from the Americas by decade from 1500 to 1660. Hamilton reports these in pesos (450 maravedís). I have converted them into metric tons of silver assuming that each maravedís is 0.094g of silver. The gray line plots TePaske's (2010) estimates of production of gold and silver in the New World by decade from 1500 to 1660. TePaske reports these in pesos (272 maravedís). I have also converted these into metric tons of silver assuming that each maravedís is 0.094g of silver.

to a flood.

Figure 4 plots estimates of Spain's imports of American treasure from 1500 to 1660 by Hamilton (1934) as well as estimates of production of gold and silver in the Americas by TePaske (2010). Most American gold and silver reached Europe through Spain over this period. The import and production measures plotted in Figure 4 differ for several reasons. First, both estimates are subject to error since sources are imperfect. Second, some gold and silver was used as money in the Americas and some was exported to Asia. These shares were small initially but grew over time. Third, Hamilton's estimates may understate imports to Spain due to illicit smuggling.

These imports of gold and silver from the New World constituted a huge monetary shock for Europe. The stock of gold and silver in Europe prior to 1492 is difficult to estimate. But a rough estimate by Glassman and Redish (1985) indicates that the combined value of gold and silver money in Europe in 1492 was equal to the value



of 3,541 metric tons of silver.<sup>1</sup> The estimates in Figure 4 indicate that around 1600 imports to Europe of gold and silver every 15-20 years were larger than the entire pre-1492 stock of monetary gold and silver on the continent. No wonder prices rose. In fact, it may seem puzzling that prices didn't rise even more. An important reason for this is that Western Europeans reexported a large amount of gold and silver to the Baltic, the Levant, and Asia to finance imports of goods from these areas. In fact a large share of all imports to Europe from Asia were financed with silver and gold rather than with the sale of European goods in Asia. For example, over 2/3 of the exports of the British East India Company to Asia was precious metals between 1660 and 1700 (Cipolla, 1976/1993, p. 220). The gold and silver that entered Europe through Spain, thus, didn't only flow to other countries in Western Europe but flowed all over the Old World.

The coincidence in the timing of the Price Revolution and the massive flow of gold and silver into Europe strongly supports the notion that the Price Revolution was caused by the inflow of gold and silver from America in the 16th and 17th centuries. In light of this, it is perhaps surprising how controversial this explanation has been. Both contemporaries and later scholars have proposed a myriad of other explanation for the sharp rise in prices in 16th and 17th century Europe. The most prominent alternative explanations pin the rise in prices on monopolists and middlemen, population increases, and high demand from abroad.

The idea that inflation is caused by monopolists and middlemen that raise prices by engrossing, forestalling, and regrating—to use Hamilton's (1934) language—is a common theme across time and space. However, this idea is flawed. Monopolists and middlemen can raise the price of the goods they sell relative to the prices of other goods. They cannot raise the overall level of all prices. Someone may object: What if there are monopolists and middlemen selling all products? Won't that make the price of all products high? But high relative to what?

Perhaps monopolists and middlemen make prices high relative to wages. This is in fact possible. But this is not what happened over the course of the Price Revolution. Both prices and wages rose by large amounts in Europe from 1500 to 1650. Such a general increase in both prices and wages cannot be explained by the nefarious actions of monopolists and middlemen.

It is true that real wages—i.e., wages relative to prices—in England fell between 1500 and 1650 (see Figure 2 in chapter XX [Malthus chapter]). Adherents of the monopolists and middlemen view can point to these movements as evidence supporting their thesis. But several objections can be raised against this argument. First,

chapter XX [Malthus chapter] provides an alternative explanation for the fall in real wages between 1500 and 1650 based on Malthusian population dynamics. Second, the fall in real wages may be interpreted as evidence that monetary expansions can lower real wages in cases where real wages are initially stuck at levels that are “too high”—i.e., above market-clearing levels. Third, it is not clear why monopolists and middlemen became so much more problematic during the period 1500-1650 than in other periods. What is really special about this period of time is the huge inflow of silver (and gold).

The idea that population increase caused the Price Revolution is also flawed. It is an application of the Malthusian ideas discussed in chapter XX [Malthus chapter]. The basic idea is that larger populations increased demand for food and thereby drove up food prices. But again, this is an argument about one price relative to another, not an argument about the general increase in all prices. As we discussed in chapter XX [Malthus chapter], increases in the population will result in lower real wages in a Malthusian economy (i.e., the price of food rising relative to wages). Population pressure can therefore explain the fall in real wages over the period 1500 to 1650, but not the large increase in the overall prices level (including nominal wages).

The idea that the Price Revolution was caused by excessive demand from abroad is less flawed. Our medieval economy model implies that increases in the money supply result in high demand in the short run. This demand does not particularly come from abroad. But contemporary complaints about excessive foreign demand may simply have been due to tribalism and political posturing. This explanation, however, leaves open the question why demand rose. The monetary theory of the Price Revolution provides an explanation for this: demand rose because of the sharp increase in the money supply.

One of the oldest surviving formulations of the quantity theory of money that underlies our medieval model is by the French scholar Jean Bodin in the context of a debate about the causes of the Price Revolution (Bodin, 1568). Bodin was responding to an explanation put forth by Sieur de Malestroit that the increase in prices experienced in France at the time was due to civil unrest, bad harvests, loss of labor after famines, and bullion exports (de Malestroit, 1566)—an early example of confusion about the causes of inflation which has been and remains very common. Even earlier discussions of the quantity theory have survived by the French scholar Oresme from around 1350, the Polish scholar Copernicus from around 1520, and members of the School of Salamanca in Spain in the 16th century. It seems likely that many prior scholars formulated some version of the quantity theory, but that records of

this have been lost.

Let's now return to Figure 3 and consider the difference between the grey line and the black line. The gray line plots the price level in pounds sterling, while the black line plots the price level in grams of silver. The fact that the gray line rises more than the black line over the course of the Price Revolution (and falls less before 1500) reflects the fact that the pound sterling was gradually debased—i.e., its metal content was reduced. From 1500 to 1650, the degree of debasement was substantial. I have normalized the two lines in the Figure 3 such that they are both equal to 100 in 1500. By 1650, the grey line has risen to 500, while the black line has only risen to roughly 300. This implies that over the 150 year period of the Price Revolution, the pound sterling was debased cumulatively by 66%. This alerts us to another reason why being on a gold standard may not guarantee price stability: the government may face strong pressure to debase the currency. We will discuss the main reasons for debasement during the middle ages and early modern period in more detail in section 8 below.

Importantly, Figure 3 also shows that debasements don't in all cases reduce price stability: the gray line is more stable prior to 1500 than is the black line. It thus seems that the English crown debased the currency between 1260 and 1500 at nearly exactly the rate needed to maintain price stability as the price of silver in terms of goods rose over this period.

Why was the price of silver rising in terms of goods? Our medieval model suggests one explanation: if the size of the economy (i.e.,  $Y^*$ ) increases faster than the supply of money in the economy, the price level will fall (i.e., the value of silver coins will rise in terms of goods). We can see this from the quantity equation in its difference form  $\Delta \log M_t = \Delta \log P_t + \Delta \log Y_t$ . If  $\Delta \log Y_t$  is larger than  $\Delta \log M_t$ , then  $\Delta \log P_t$  must be negative. (Increases in  $Y_t$  can be thought of as being driven by increases in  $Y^*$ .)

## 7 Ghost Money

Today the unit of account in most countries is a piece of money that is “real” in the sense that there exists a coin or a bill that corresponds to this unit of account. For example, in the United States, the unit of account is the U.S. dollar and actual U.S. dollar bills circulate in the U.S. economy. In Japan, prices and wages are quoted in yen and one can actually pay for things with yen coins. Given the tight link between

the units of account that exist today and actual coins and bills, it is tempting to think that this is how things must be, i.e., that the unit of account is bound to correspond to the physical money circulating in a particular economy.

But this is not at all so. In fact, historically, the unit of account has often deviated from the coins and bills circulating in economies in ways that at first may appear quite strange. Cipolla (1956) gives a classic account of this:

For centuries the people of Western Europe went on using monetary terms such as *libra* (pound), *solidus* (shilling), *denarius* (penny), and *florinus* (florin) in fixing the prices of their commodities and services, in establishing any sort of debts, or in keeping accounts. Everybody used those terms and carried on transactions in them ... Yet, and here the mystery begins, during the greatest part of the Middle Ages and the first centuries of the modern period, with the exception of a few short periods, nobody ever saw many of those "moneys" about which everybody talked. For instance, nobody for centuries ever saw a real pound, for the simple, but paradoxical, reason that the pound during the greatest part of its life did not materialize into a real, visible, and touchable coin. It was a ghost money. (page 38)

An important lesson from monetary history is that the unit of account can live a life of its own quite independent of the coins and bills circulating in the economy (and other mediums of exchange). The history of units of account in Western Europe over the past 2000 years illustrates this clearly.

We start this story with the monetary system of the late Roman Empire. From the fourth century C.E., this was based on the *solidus*, a gold coin that weighed roughly 4.5 grams. The *solidus* continued to be minted by the Byzantine Empire after the fall of Rome and was the dominant coin used in Mediterranean trade until the end of the 7th century. The Muslim conquests of the 7th and 8th century substantially altered the balance of power in the Mediterranean and gave rise to a rival coin: the muslim *dinar*. The *dinar* was the result of a monetary reform instituted by Calif Abd Al-Malik ibn Marwan in the 690s. From that time until the thirteenth century, the *solidus* (also called the *nomisma*) and the *dinar* were together dominant coins for international trade in the Mediterranean.

The disintegration of the Western Roman Empire in the 5th century C.E. resulted in significant regress in monetary matters in Western Europe for some time. Coins became scarce. Obligations were often settled in kind rather than with coins. The

units of account continued to be concepts such as *solidi* and *denarii*, which corresponded to coins that existed or had existed. But the medium of exchange in most transactions were everyday commodities (or labor). In this sense, the economy shifted closer to a barter economy, but one with coin-based units of account.

An important change occurred in the eight century when the Carolingian king Pepin the Short reformed the monetary system of the Carolingian Empire (soon to become the Holy Roman Empire). His son Charlemagne finalized these reforms in the 790s. The new system—which became the dominant monetary system in Western Europe—was based on a silver penny (*denarius* in Latin or *denier* in French) weighing about 1.7 grams. This penny was the only coin minted in Western Europe for several hundred years.

The most prevalent unit of account prior to these reforms was the *solidus* (*shilling* in English or *sou* in French). The value of a *solidus* at the time was roughly equal to twelve of the new pennies. To link the new system with the old system, the Carolingian reform declared a *solidus* (*shilling*) to be equal to twelve pennies. Finally, the weight of the penny was determined such that 240 pennies could be struck from a pound (*libra* in Latin or *livre* in French) of silver.

The Carolingian monetary system was therefore based on three units:

$$1 \text{ pound} = 20 \text{ shillings} = 240 \text{ pennies.}$$

Only one of these three units corresponded to a real coin (the penny). The *shilling* was the name of an old coin used to link the old and the new monetary systems. The *pound* was not a coin either. It was originally a unit of weight of silver. This somewhat odd monetary system consisting on dozens and scores of dozens was remarkably long lived. It was only in 1795 that France replaced this system with a decimal system (1 franc equal to 100 centime). The last country to use this monetary system was Britain, which only switched to a decimal system in 1971.

The Carolingian penny was a relatively modest monetary unit. It was a convenient unit for many everyday transactions. But for larger transaction such as wholesale merchant transactions, it was a small unit. Merchants therefore found it useful to record trades in pounds and shillings. Cipolla (1956, p. 41) discusses how initially pounds referred to the weight of the silver involved in a transaction, but over time pounds came to mean simply “240 pennies.” Similarly, *solidus* (*shilling*) referred initially to the gold coin of that name, but over time came to mean simply “12 pennies.”

This distinction was not material at first, but for reasons discussed in detail in

section 8, the penny was progressively debased over time, i.e., its silver content was reduced. This process meant that as time passed the amount of silver in 240 pennies fell below 1 pound by greater and greater amounts. At that point there arose a conflict between the monetary unit “1 pound” and the pound as a measure of weight. It is not obvious what should happen in such a circumstance. Some might think it more natural that the monetary unit “1 pound” remain equal to 1 pound of silver and that its exchange rate with the penny should simply have varied as the penny was debased. But this is not what happened. The pound as a unit of account came to live a life of its own independent of its original meaning as the weight of some number of pennies.

The same is true of the solidus (shilling). Originally, the gold solidus coin was equal in value to 12 pennies. As pennies were debased, the gold solidus coin came to be worth more than 12 pennies. However, the solidus as a unit of account remained 12 pennies. In monetary matters, it was therefore important to distinguish between the unit of account “1 solidus” and the actual gold solidus coin, which was worth more than 12 pennies. As with the pound, the ghost money had begun to live a life of its own.

In the 12th and 13th centuries, various states in Western Europe began minting new coins representing denominations larger than the penny. The florin minted in Florence from 1252 onward and later the ducat minted in Venice became important coins in Mediterranean trade. The gold florin was originally equivalent to 240 local pennies. It therefore materialized the local pound. In France Louis IX issues a gold half-pound coin (initially worth 120 pennies) as well as a silver shilling worth 12 pennies. However, “that happy moment when the ghosts materialized and let themselves be touched and seen by everybody did not last for a long time” Cipolla (1956, p. 43) explains. Generally, the penny was debased more than the larger gold coins resulting in the value of the gold coins rising relative to the penny as time passed. As this happened, “the ghosts left their material features and began again their mysterious, invisible life” (ibid).

The story of the Milanese florin as told by Cipolla (1956, p. 46-47) drives the point home. The Milanese florin was rated at 120 pennies when it was first struck in 1252. However, as the penny was debased its value rose. By 1340 it was worth 384 pennies. At that point, the debasement of the penny seized for some time and the exchange rate between the florin of Milan and the local penny remained stable for about 60 years. Over this period, the people of Milan came to use the florin as an additional unit of account. The system of denominations in Milan towards the end

of the 14th century therefore had four denominations:

$$\begin{aligned}1 \text{ florin} &= 1.6 \text{ pounds} = 32 \text{ shillings} = 384 \text{ pennies} \\1 \text{ pound} &= 20 \text{ shillings} = 240 \text{ pennies} \\1 \text{ shilling} &= 12 \text{ pennies}\end{aligned}$$

Two of these units were ghosts (the pound and the shilling) and two were actual coins (the florin and the penny). In the fifteenth century, the debasement of the penny relative to the florin resumed. But by this point, the practice of using the florin as a unit of account was sufficiently ingrained—for example in accounting practices—in Milan that the Milanese simply continued to use the term “florin” to mean 384 pennies, just as they used the term “pound” to refer to 240 pennies even though 240 pennies were no longer equal to 1 pound of silver. The florin as a unit of account became a ghost, i.e., an abstract monetary unit. The value of the actual gold florin coin rose over time relative to this abstract florin. By 1445, a gold florin coin was worth about 768 pennies. In other words, 1 “real” gold florin was worth 2 “ghost” florins.

Cipolla (1956) gives several other examples. The Genoese came to use the term “florin” as an abstract unit of account equal to 300 pennies (25 shillings) even as the Genoese gold florin became worth more than this. The Venicians came to use the term “ducat” similarly to mean 124 shillings irrespective of the value of the actual gold Venician ducat. It was universally the case that the term shilling meant 12 and pound meant 240. But florin meant different things in different cities, as did other units.

This discussion suggests that contemporary merchants (as well as later day scholars) had to contend with a rather complicated monetary system during the Middle Ages. It seems likely that all this complexity was costly in various ways and therefore detrimental to trade: those not sufficiently knowledgeable could be tricked; contracts might be open to several interpretations unless monetary matters were spelled out in great detail (e.g., did the term “florin” as used in the contract mean the gold coin of that name or the abstract monetary unit). What prevented the monetary authorities of this era from instituting a simpler and more transparent monetary system? We take up this question in the next section.

## 8 The Big Problem of Small Change

During the decades prior to World War I, most countries in the world were on a gold standard. Many accounts of monetary history begin with this classical gold standard period, proceed to discuss its downfall, and then detail the tumultuous history of monetary arrangements over the 20th century. This narrative can give the false impression that all of history prior to 1914 was characterized by a gold standard. In fact, despite its great fame, the classical gold standard period was very brief: it lasted only from the 1870s until 1914.

In chapter XX, we discuss the reasons for the downfall of the gold standard. Here, however, we consider why the gold standard didn't emerge earlier. A central challenge faced by monetary authorities for many centuries—and arguably the reason why the gold standard did not emerge earlier—was how to maintain coins of different value in circulation simultaneously. Recall that prior to bills, bank deposits, and other modern media of exchange, coins were the main medium of exchange in society. Coins of vastly different value were needed for the simple reason that the scale of transactions varied enormously. On the one hand, low-value coins were needed for everyday household transactions. On the other hand, high-value coins were needed for large-scale transactions such as wholesale transactions of merchants. As we will see below, monetary authorities faced particularly severe problems in providing for low-value coins—providing for small change. This challenge is sometimes referred to as the “big problem of small change” (Cipolla, 1956; Sargent and Velde, 2002).

Practical considerations implied upper and lower limits on the size of coins. On the one hand, very large coins were too easy to counterfeit, for example, by filling the interior with a base metal (Redish, 2000, p. 19). Bodin (1606/1962, p. 689) argued that 31g was a rough upper limit. On the other hand, coins could not be so small that they were easily lost or difficult to pick up and count. Bodin argued that coins smaller than 1.3g were difficult to stamp and too brittle. The smallest denomination English coins in the 15th century, where actually quite a bit smaller than this. The farthing (1/4 of a penny) weighed only about 0.25g at this time (Redish, 2000, p. 109). For comparison, a U.S. dime is 2.27g and a euro 1 cent coin is 2.3g.

Clearly, the farthing, the halfpenny, and even the penny were extremely small coins in the 15th century. Yet, the value of these coins was not at all trivial from the point of view of everyday transactions. Allen (2001) estimates that the day wage of building laborers in London in 1425 was 5d (i.e., 5 pence). A penny was therefore



20% of a day's wages for a building laborer and the tiny farthing was 5% of a day's wages for such a worker. For comparison consider a worker in 2020 who made \$15 per hour and worked an 8 hour day. 20% of such a worker's daily wages was \$24, while 5% was \$6. This means that the smallest denomination silver coin in England in the 15th century was to a building laborer of that time roughly comparable to a \$5 bill in 2020.

Conducting everyday transactions with a \$5 bill as the smallest denomination seems cumbersome enough. But the situation was actually worse than this since there was a chronic shortage of farthings and halfpennies. Minting such tiny coins was expensive (relative to their modest value). While sovereigns may have paid some attention to the public service aspect of keeping small coins in circulation, they also put considerable weight on the profits minting generated. For this or some other reason, they failed to enact policies that resulted in adequate supply of farthings and halfpennies (Redish, 2000, p. 108-115). In many cases, therefore it is likely that the smallest coins available in the 15th century in England were pennies. Imagine how cumbersome everyday transactions would be if the smallest denomination of money was a \$24 coin (at 2020 prices).

The fundamental difficulty monetary authorities faced in the 15th century in providing an efficient medium of exchange was that "silver—an object of intermediate value—was too valuable to provide a medium of exchange for the smallest transactions and too cheap to use for high-value transactions" as economic historian Angela Redish put it in her excellent book on bimetallism which this section draws on heavily (Redish, 2000, p. 19). One solution for small transactions was to mint coins of varying fineness, i.e., coins made of an alloy of silver and copper with a smaller share of silver than "sterling silver" coins which contained 92.5% silver (or at some times 91.67%). Such coins were called "billon" in France and "vellon" in Spain. Billon coins were never minted in England.

An important drawback of billon coins was that it was difficult—even for skilled assay-masters—to assess their fineness (Redish, 2000, p 21-24). This resulted in strong incentives to counterfeit billon coins, i.e. mint replicas of official coins with lower silver content than the official coins. Also, different polities on the continent of Europe would mint very similar deniers (pennies) and would each have an incentive to lower slightly the proportion of silver in the deniers they minted relative to existing deniers in circulation. This would allow them to offer better prices for the coins they minted than competing polities. Over time, this led to very substantial debasement of the denier on the continent. England was spared this dynamic by

being an insulated island economy.

Silver coins were also problematic for large value transactions—such as wholesale merchant transactions—due to the fact that their value was modest per unit weight. Gold was between 10 and 15 times more valuable by weight throughout the period 1000 to 1800. (The difference was even larger by volume due to gold’s high density.) This meant that gold coins were much preferred for merchant trade. As we discussed in section 7, Florence, Genoa, Venice and other city states started to mint gold coins in the 13th century to cater to demand from merchants.

A monetary system based on a combination of silver and gold coins had the important benefit that it could supply coins with a wide range of value. For this reason, the practice of minting both silver and gold coins spread throughout Western Europe after the 13th century. In other words, the monetary system of most states in Western Europe evolved toward a *bimetallic standard*: a monetary system with coins made of two precious metals (the mixed-in copper didn’t count). Unfortunately, bimetallicism had a very significant drawback: it proved quite difficult to maintain coins made from silver and gold in circulation simultaneously at constant relative value.

This difficulty is best explained by example. In the early 17th century, English authorities minted a gold coin called the “unite” which was legal tender for 20 shillings (i.e., 1 pound). This coin contained 0.2689oz (8.36g) of gold, which means that unites were minted at 74.4 shillings per ounce of gold.<sup>2</sup> At the same time, the English authorities also minted the silver “sixpence” (1/2 shilling). The sixpence contained 0.089oz (2.78g) of silver. This means that the sixpence was minted at 5.6 shillings per ounce of silver. The ratio of the weight of a silver shilling and a gold shilling was therefore  $74.4/5.6 = 13.3$ .

Gold and silver traded on world commodity markets and the relative price of one in terms of the other varied over time. In the 17th century, for example, the price of gold in terms of silver on world commodity markets was gradually rising (due to the huge amounts of silver entering Europe from the Americas). When it rose above 13.3, gold became “undervalued” at the English Mint (and silver “overvalued”). Gold was undervalued at the Mint in the sense that it was more profitable for someone who owned gold to sell the gold for silver in the world market and take the silver to the Mint than it was to take the gold directly to the Mint.

Suppose for concreteness, that the world market price of gold in terms of silver was 14, and for simplicity that the English Mint did not charge for minting coins. In this case, a person who owned an ounce of gold could either take the gold to the

Mint and receive 74.4 shillings or sell the ounce of gold for 14 ounces of silver on the world market, take this silver to the Mint and receive  $14 \times 5.6 = 78.4$  shillings.

In this situation, no one would mint gold coins. If the difference between the price of gold in terms of silver at the Mint and in world markets was large enough, it might even become profitable to melt down gold coins, sell the gold in world markets for silver, and take that silver to the Mint. Such a situation would eventually lead to only silver coins circulating with the adverse consequences for trade that this entails.

In reality, transaction costs meant that gold and silver coins could circulate along side each other even when the gold-silver price ratio at the Mint differed from the gold-silver price ratio in the world market by a non-trivial amount. One type of transaction cost was the fee the Mint charged, i.e., the gross seignorage. For example, a person presenting gold to the English Mint in 1615 received unite coins worth only 72 shillings per ounce even though 74.4 shillings worth of such coins were minted per ounce. The Mint kept the remaining 2.4 shillings per ounce as seignorage. (The Mint also charged seignorage for minting silver coins.) In addition to seignorage, melting coins, shipping gold abroad for sale, and reimporting silver entailed significant costs. These costs implied that arbitraging differences in the gold-silver price ratio at the Mint versus in world commodity markets was only profitable when this difference was quite substantial.

A second problem that led to undervalued coins was wear from circulation and clipping. Normal use of a coin leads the coin to gradually lose some of its weight. For gold and silver coins, this can be substantial over longer periods of time. In addition, "clippers" would profit from clipping small amounts of silver off of coins, collecting these clippings, melting them down, and selling them as bullion.

Wear from circulation and clipping implied that older coins were lighter than newly issued coins. If older coins continued to circulate "by tale" (i.e., by number) as opposed to by weight, this meant that newly issued coins were undervalued at the Mint relative to the older coins. For everyday household transactions, weighing coins every time they transact would be very cumbersome. In practice, therefore low value coins generally circulated by tale. But this meant that people had an incentive to pay with older lighter coins rather than newer heavier coins. Heavier coins tended to be exported or melted down. This is another example of Gresham's Law: bad money drives out good money. In this case, the worn out (lighter) coins are "bad money", while the newer (heavier) coins are "good money".

Over time, the difference in weight between older worn out coins and newly

issued coins would cause a shortage of official coins and a proliferation of counterfeit coins as well as foreign imitation coins. These counterfeit and foreign imitation coins became cheaper to produce than official coins as the stock of coins in circulation became more and more worn (and therefore lighter). The price of silver in the world commodity market would rise above what people could get by bringing silver to the Royal Mint. Once this had happened, very few new official coins would be minted unless the monetary authorities were willing to subsidize the minting of additional official coins. This whole process would lead the stock of coins in circulation to become more and more heterogeneous, which risked undermining the efficiency of the coins as a medium of exchange. As a result, complaints about the state of the coinage would multiply (Feavearyear, 1963, p. 10-20).

These problems were arguably a primary reason for debasement of low value coins in England and other Western European countries after the 14th century. Faced with a stock of coins in greater and greater disarray, monetary authorities would periodically choose to realign the silver content of newly issued coins with those in circulation by reducing the weight of newly issued coins, i.e., debase the currency. This would encourage people to bring silver to the Mint and discourage counterfeiting. The difference between the black line and the gray line in Figure 3 reflects this process of gradual debasement of the currency. Evidently, wear and clipping of coins was another reason—in addition to increases in the supply of silver—why being on a commodity standard did not guarantee price stability in the late medieval and early modern period.

Sometimes coins would be debased not for the reasons discussed above, but to raise revenue. Debasement allowed states to raise their seignorage rates temporarily and also increase the quantity of coins minted (older coins often lost their legal tender status). For both of these reasons, debasements were a source of revenue for the state. Such debasement was however unpopular. States therefore refrained from such debasements except when their fiscal needs were dire (e.g., during wars). Fiscal debasements were also less frequent in states with stronger institutions, where fiscal needs could be met in other ways.

There are also instances when the states decided to recall coins in circulation and recoin without depreciating the currency. One instance of this is the English recoinage of 1696. In such a case, the holders of the worn coins bore the cost. They had taken a light coin as payment for (say) a shilling when its true silver content was only 9d (9 pence). They must then return the coin to the Mint in exchange for 9d worth of newly issued coins. In the case of the recoinage of 1696, the

English Exchequer (treasury) offered to purchase worn and clipped coins prior to the recoinage and therefore bore most of the fiscal cost of the recoinage.

Today, our monetary system no longer suffers from the “big problem of small change.” The reason for this is that an effective solution to this problem was eventually devised. The solution was to issue *token coins*: coins that have a monetary value above the value of the metal they are made of. This system is now universally used by monetary authorities to supply low value coins. But despite its simplicity, it took centuries to work out. It was only in 1816 that England first adopted this system, and other nations adopted it still later (e.g., the United States in 1853 and France in 1865).

Why did it take so long for the big problem of small change to be solved? One reason was surely the widespread metallist fallacy that money must be backed by an object that has non-monetary value (such as gold or silver). Our discussion of stone money on the island of Yap earlier in the chapter, not to mention the near universal use of fiat money for the last 50 years has clearly demonstrated that money need not be backed by items with non-monetary value. But this belief was clearly widely held until not so long ago and may have clouded the thinking of monetary authorities.

Another reason emphasized by Redish (1990, 2000) was technological. The fact that the face value of a token coin was higher than the value of the metal the coin was made of created potential for profitable counterfeiting. For example, if 5s 6d (5 shillings and 6 pence) were coined from an ounce of silver when the price of silver on the market was only 5s, counterfeiters could earn a profit as long as their costs were lower than 6d per ounce. (Their profits would be 6d minus their cost per ounce.) If counterfeiting was not very costly, this would place a tight bound on the difference between the face value of a token coin and the value of the metal it was made of.

The traditional technology for making coins was by hammering blanks: a blank was placed between two dies and the top was struck several times with a hammer. This method created coins of uneven size and quality. Such coins were easily counterfeited. In the 17th century, this method was replaced by a fly-press and the edges of the coins were milled or engraved. The fly-press created somewhat more uniform coins. But the major advantage of this method was that the milled or engraved edges made clippings more obvious and therefore discouraged clipping of coins.

Redish (1990, 2000) argues that it was not until the advent of steam power in the early 19th century that coins could be made with sufficient uniformity and quality to substantially raise the costs of counterfeiting. Steam power was first adapted to minting by Matthew Boulton (James Watt’s partner). Between 1805 and 1811, the

Royal Mint was fitted with Boulton's machines. In addition to producing coins of superior quality, Boulton's machines had the advantage of being an expensive proprietary technology that was hard for counterfeiters to replicate. These technological changes removed a crucial impediment to the adoption of token coins.

Ever since Sir Isaac Newton, as Master of the Mint, overvalued gold coins in England in 1717, very little silver coinage had been minted in England. This led to a severe shortage of small denomination coins throughout the 18th century. The silver shillings and sixpence that did circulate became extremely worn as the century wore on. By 1798, shillings in circulation had lost 25% of their weight and sixpence 38% (Redish, 2000, p. 142). This sorry state of affairs eventually led to reform of the silver coinage in England. The Coinage Act of 1816 called for new silver coins to replace the old. Significantly, these new silver coins would be token coins: 5s 6d would be minted from an ounce of silver, whereas the market price of silver at the time was 5s. The Act ended "free coinage" of silver—i.e., the notion that anyone could bring silver to the Mint for coining. This meant that Britain had shifted from an official bimetallic standard to a gold standard.

Ending free coinage of silver was crucial. Since the face value of token coins is higher than the metal they are made of, it is profitable to create them. This means that their supply must be controlled. If anyone had been allowed to exchange silver bullion for such coins in arbitrary quantity the Mint would have been flooded with silver and the price of the coins would eventually have been driven down to the value of the metal they were made of. Reflecting the rather foggy understanding the English authorities had of the reform they were undertaking, it was somewhat by lucky happenstance that the Coinage Act did not include a clause allowing for free coinage of silver (Redish, 2000, p. 148-149).

Another crucial feature of a successful monetary system based on token coins is that the monetary authority must commit to redeem the token coins at face value. Without this guarantee, the value of the token coins will not necessarily remain equal to their face value and, in particular, might well fall towards the value of the metal they are made of. The importance of this type of convertibility was another aspect of the reform that seems not to have been well understood by the English authorities initially. In this case, it was the Bank of England—at the time a private for-profit institution—that helped make the new token silver coins a success by guaranteeing their convertibility into gold coins at par. Eventually, a confrontation between the Bank and the Exchequer forced the Exchequer to take on the responsibility of guaranteeing convertibility.

The coinage system Britain adopted in 1816 proved to be the solution to all the problems discussed in this section. It constitutes a major advance in monetary practice and was eventually universally adopted. The crucial element of the “standard formula” Britain devised was the issue of token subsidiary coins that were freely convertible at par and of sufficiently high quality that counterfeiting was not profitable. This system allowed monetary authorities to issue coins with a wide range of values without difficulty. Actually, this system allowed monetary authorities to completely stop thinking about the composition of the money supply across denominations. (The Fed doesn’t care how much of the money supply is nickels versus quarters.) With the big problem of small change solved, the focus of monetary policy shifted from providing an efficient medium of exchange to maintaining stable prices.

## 9 The Crime of 1873

In the U.S. presidential election of 1896, William Jennings Bryan was the candidate of the Democratic party. The U.S. was on the gold standard by this point. It had undergone the transition from a traditional bimetallic standard with its “big problem of small change” to the standard formula with token low value coins discussed in section 8 earlier in the 19th century. However, Bryan’s principle campaign issue was “free silver” at “16 to 1”—that is, a return to a bimetallic standard. By 1896, the Coinage Act of 1873, which demonetized silver—and thereby shifted the U.S. from a bimetallic standard to a gold standard—was by many considered a “grave wrong,” a “conspiracy,” a “great legislative fraud,” or simply “the crime of 1873” (Barnett, 1964; Friedman, 1990b).

William McKinley was Bryan’s Republican opponent. He ran on a platform of remaining on the gold standard. Even though Bryan ran a memorable campaign that included one of the most famous convention speeches of all time in which he thundered “You shall not press down upon the brow of labor this crown of thorns, you shall not crucify mankind upon a cross of gold”, Bryan lost the election by a vote margin of 4.3 percent (and an electoral college margin of 271 to 176).

Given all the problems associate with being on a bimetallic standard discussed in section 8, it may seem surprising that Bryan’s central campaign issue was the return to a bimetallic standard. Also, why did Bryan argue that being on a gold standard was a “crown of thorns” for labor? Furthermore, why were technical matters of

monetary policy so central in a presidential election campaign during the classical gold standard period? Isn't that period held up as a golden era of monetary stability? Why all this agitation about monetary matters during that era?

Answering these questions is not merely of historical interest. It helps us understand key challenges associated with being on a gold standard (or any specie standard, for that matter). As in section 6, we will see that when countries are on a specie standard their macroeconomic outcomes are at the mercy of unpredictable variation in the supply of bullion.

The Coinage Act of 1792 defined the dollar as the basic monetary unit of the United States. According to the act, the dollar was equal to 371.25 grains (24.06g) of pure silver or 24.75 grains (1.60g) of pure gold. The act authorized *free coinage* of both gold and silver. Coinage was free in two senses of the word: first, anyone could take bullion to the mint in unlimited amounts and receive coins in exchange; second, coinage was free of charge, i.e., the mint did not charge seignorage.

The Coinage Act of 1792 thus placed the United States on a bimetallic standard with a gold-silver price ratio of 15 to 1. The principle architect of the Act was treasury secretary Alexander Hamilton. Hamilton had recommended the ratio of 15 to 1 because this was the prevailing market rate at the time the act was being drafted. Hamilton's hope was that both gold and silver coins would circulate in the U.S. and the U.S. economy would thus benefit from an efficient medium of exchange for transactions over a wide range of values (Redish, 2000, p. 212-213).

Hamilton's hopes in this regard were thwarted by the fact that the market price of gold in terms of silver rose shortly after 1792 and remained above 15 thereafter. We can see this in Figure 5, which plots the price of gold in terms of silver from 1690 to 1870. The fact that the market price of gold in terms of silver was above 15 after 1792 meant that gold was undervalued at the U.S. mint and silver was overvalued. The public therefore brought silver to the U.S. Mint rather than gold and the U.S. was effectively on silver standard.

In the early 19th century, there was considerable debate in the U.S. about the monetary standard. Some argued that the absence of gold coins was detrimental to commerce, while others argued that the use of bank notes—which was more common in the U.S. than in other countries at the time—was an effective substitute for high denomination coins. Several committees recommended raising the gold-silver price ratio at the mint to a value closer to the prevailing market price of gold with most discussion centering on a value of 15.625. In 1834, legislation was introduced to raise the gold-silver price ratio to 15.625. Shortly before the legislation passed,



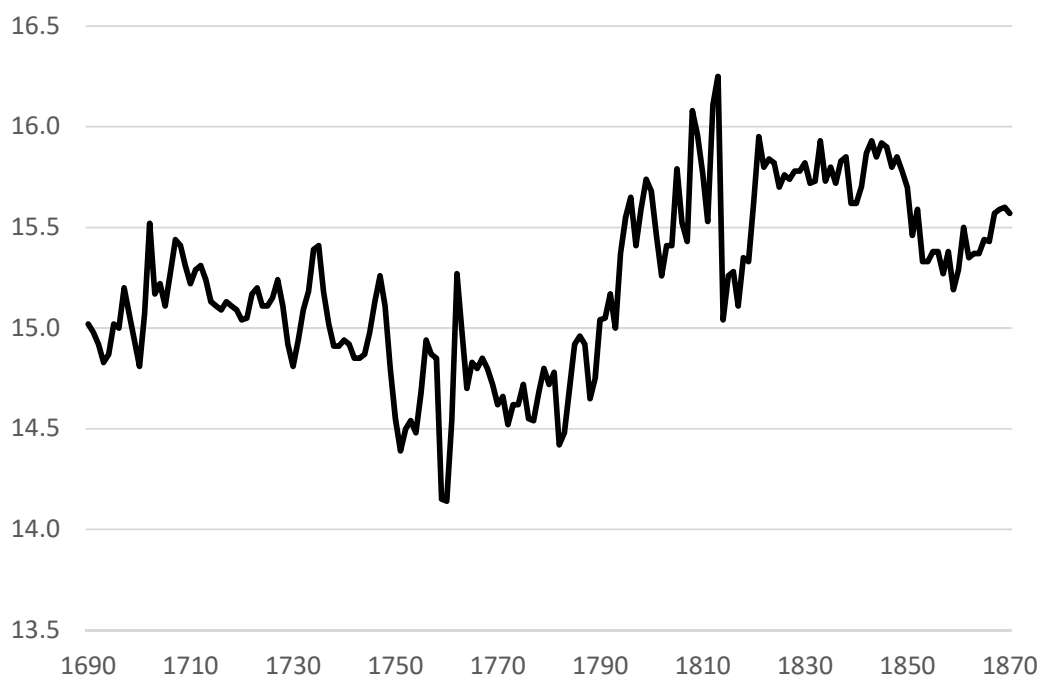


Figure 5: Gold-Silver Price Ratio, 1690-1870

*Note:* The figure plots the gold-silver price ratio (ounces of silver for an ounce of gold). The source is Officer and Williamson (2021). The construction of this series is described in Officer (Undated-b).

however, the price ratio was changed to 16. Why this was done is not entirely clear, but it was likely due to political considerations. President Andrew Jackson was at the time engaged in a battle to eliminate the Second Bank of the United States. Some think that raising the price ratio to 16 was done to undermine the Bank by encouraging the minting of gold coins as a substitute for bank notes issued by the Bank. Others have argued that the higher price ratio was chosen to benefit gold producing states (Friedman, 1990, Redish, 2000, p. 216-224).

The passage of the Coinage Act of 1834 meant that silver became undervalued at the Mint and gold overvalued. At first, the degree of undervaluation of silver was small (see Figure 5). During this period silver continued to circulate and some additional silver coins were minted (although much more gold coins were minted). Following the California gold rush (which began in 1848), the price of gold fell and silver became more substantially undervalued at the Mint. This led silver coins to be withdrawn from circulation (exported or melted down)—an example of Gresham’s Law and the “big problem of small change” (Redish, 2000, p. 227-232).

Congress in the U.S. responded to the growing shortage of small denomination coins by passing legislation in 1853 that reduced the silver content of subsidiary

silver coins (those with a face value less than \$1). This change made subsidiary silver coins in the U.S. into token coins and thus took an important step towards eliminating the big problem of small change. However, the 1853 legislation did not require the Treasury to redeem these coins and thus did not fully anchor their market value at their face value. A commitment to redeem subsidiary silver coins did not become law in the U.S. until 1879.

The outbreak of the Civil War in 1861 resulted in an enormous increase in government spending. Financing this spending was a huge challenge. One of the responses of the Union government to this challenge was to issue paper currency—so called “greenbacks.” The Union government printed large quantities of these greenbacks and used them to pay for the war effort. Greenbacks were not backed by gold or silver and no explicit promise was made to redeem them in gold or silver. Gold and silver coins continued to circulate as well, but not at a 1-to-1 exchange rate. The Union economy, thus, effectively had a dual monetary system during the Civil War with both greenbacks and gold coins serving as money at a floating exchange rate. Greenbacks were the dominant form of money in most of the country, but gold coins were also widely used for certain purposes, particularly on the west coast.

As the war progressed and more and more greenbacks were issued, their value fell. In other words, there was a large amount of inflation. Figure 6 plots the evolution of the price level in the U.S. between 1825 and 1914 with the price level in 1860 normalized to 100. The figure shows that over the course of the War, the price level (in greenbacks) roughly doubled.

After the war, there was a widespread desire in the financial community (and generally among those in the country with political power) to return to a specie standard—a system of “sound money.” This was accomplished with a sequence of acts of Congress in the 1870s. An important consideration was at what exchange rate greenback dollars should be made convertible into gold. Adopting a gold standard at the pre-War exchange rate entailed reducing the price level by a very large amount—i.e., raising the price of the greenback dollar up to the price of the old gold dollar by reversing much of the inflation of the war years. The alternative was to fix the dollar at a lower exchange rate (fewer dollars per ounce of gold). We will discuss the economics of this choice in more detail in chapter XX. The final decision—embodied in the Resumption Act of 1875—was to adopt the pre-War exchange rate. Resumption was scheduled for January 1, 1879 and was successfully executed at that time.

Our focus in this section is, however, the Coinage Act of 1873. This act listed the

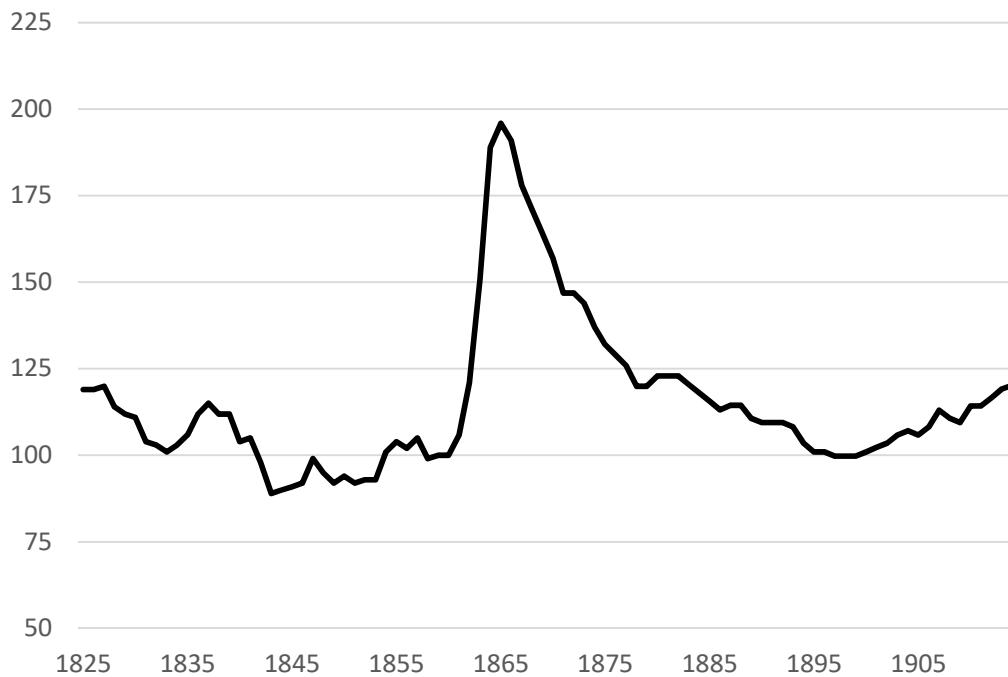


Figure 6: An Estimate of the Consumer Price Index in the U.S., 1825-1914

*Note:* The figure plots an estimate of the evolution of the consumer price index in the United States from 1825 to 1914. The price level in 1860 is normalized to 100. The source is Williamson (2021). The construction of this series is described in detail in Officer (Undated-a). It uses series constructed by David and Solar (1977), Rees (1961), Douglas (1930), as well as the Bureau of Labor Statistics.

coins to be minted in the U.S. going forward. It included gold coins and token silver coins. However, it ended free coinage of the silver dollar. This effectively moved the U.S. from a bimetallic standard to a gold standard, just as the British Coinage Act of 1816 had done in Britain. The adoption of token subsidiary silver coins had removed the traditional rationale for being on a bimetallic standard. So, the U.S. adopted Britain’s “standard formula.” The Act was passed by an overwhelming majority in both houses of Congress and attracted little attention at the time it was passed.

At roughly the same time (i.e., in the 1870s), Germany switched from a silver-based currency to a gold standard, as did France. France had operated a bimetallic monetary system with a gold-silver price ratio of 15.5 from 1803. This likely played an important role in stabilizing the gold-silver price ratio despite substantial gold discoveries in California and Australia around 1850 (Friedman, 1990a). The near simultaneous demonitization of silver in France, Germany, and the United States in the 1870s constituted a huge negative shock to the demand for silver. As a result, the price of silver began to fall relative to gold.



Figure 7: Gold-Silver Price Ratio, 1800-1914

*Note:* The figure plots the gold-silver price ratio (ounces of silver for an ounce of gold). The source is Officer and Williamson (2021). The construction of this series is described in Officer (Undated-b).

Figure 7 plots the price of gold in terms of silver between 1800 and 1914. The price ratio was very stable between 15 and 16 from 1800 until the 1870s. During the 1870s, it started to rise. By 1880 it had risen to 18 and in the 1880s it rose to 22. This increase in the price of gold in terms of silver implied that, if not for the Coinage Act of 1873, the U.S. would have resumed convertibility to specie in 1879 on a de facto silver standard as opposed to a gold standard: silver would have been overvalued at the mint and gold undervalued leading the public to bring silver to the mint rather than gold.

Why does this matter? It matters primarily because of its effect on the price level. Countries on the gold standard experienced a persistent, large deflation (fall in the price level) during the 1870s, 1880s, and early 1890s. Figure 8 plots the price level in the U.S. and Britain from 1870 to 1914. Between 1875 and 1895, the price level in both countries trended downward. Cumulatively, the price level fell by 23% in the U.S. and 18% in Britain over this period. In the mid-1890s, however, the price level in both the U.S. and Britain stopped falling and then began to rise. From that point until the beginning of World War I, these countries (and other countries on the gold standard) experienced persistent inflation rather than persistent deflation.

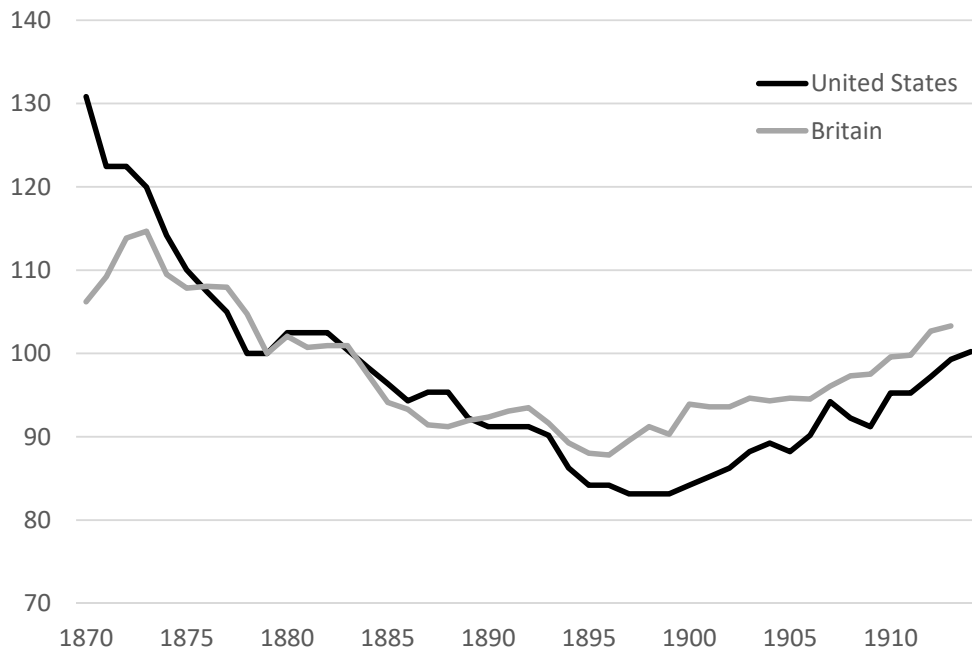


Figure 8: An Estimate of the Consumer Price Index in the U.S. and Britain, 1870-1914

*Note:* The figure plots an estimate of the evolution of the consumer price index in the United States and Britain from 1870 to 1914 (1913 in the case of Britain). Both series are normalized to 100 in 1879. The sources are the same as for Figures 3 and 6.

For the United States, the deflation between 1875 and 1895 was a continuation of a much larger deflation that began with the end of the Civil War (see Figure 6). As we discussed above, the U.S. decided to adopt a gold standard at the pre-Civil War exchange rate between the dollar and gold. This entailed raising the value of the greenback dollar—equivalently, reducing the price level—by a very large amount. The large deflation between 1865 and the mid-1870s was primarily caused by this policy.

In contrast, the deflation between 1875 and 1895 was a global phenomenon affecting all gold standard countries. Its cause likely lies in the supply and demand for gold. In particular, the demand for gold increased substantially over this period for two reasons. First, countries representing a substantial fraction of world GDP adopted the gold standard and, thus, demanded large quantities of gold to back their money supply. (The money supply was rapidly switching from metal coins to paper notes during this period. We will discuss this in more detail in chapter XX [Money and Banking Chapter].) Second, the world economy expanded at a rapid clip during this period as millions of Europeans emigrated to the New World and

the Industrial Revolution spread to larger and larger parts of the world.

To understand the effect that these developments had on the price level in gold standard countries, it is useful to refer back to the model that we developed in sections 4 and 5. The quantity equation implies that

$$\Delta \log M_t = \Delta \log P_t + \Delta \log Y_t. \quad (10)$$

(See discussion on page 26.) Rearranging this equation, we get

$$\Delta \log P_t = \Delta \log M_t - \Delta \log Y_t. \quad (11)$$

The logic for this equation is that changes in the price level are determined by the balance between two factors: 1) changes in the supply of money  $\Delta \log M_t$ , which are heavily influenced by changes in the quantity of gold (i.e., gold mining), and 2) changes in the demand for money, which are determined in our simple model by  $\Delta \log Y_t$ . If the supply of money increases faster than the demand for money, prices will rise. If, however, the demand for money increases faster than the supply of money, prices will fall. During the period 1875 to 1895, mining of gold—while considerable—did not keep up with the rapid increase in the demand for gold for monetary purposes. As a result, the price level in countries on the gold standard fell persistently.

You may again ask: Why does this matter? Why, in particular, did the public get sufficiently exercised about this deflation for it to be a major campaign issue in a presidential election? One reason is that changes in the price level affect the real value of debt contracts. Debt contracts are typically written in nominal terms (e.g., in dollars). When the price level falls, the real value of debts rises. This is good for creditors, but bad for debtors. Debtors, therefore, typically do not like deflation, while creditors do not like inflation (which erodes the real value of debt contracts).

This idea is perhaps best grasped through an example. Consider a farmer who takes out a \$1,000 loan (a considerable sum in the 1870s) to build a barn, buy livestock, or otherwise improve his or her farm. Suppose for simplicity that the farmer must pay 5% in interest per year and repay the principle at the end of a 5 year term. If the price level falls by (say) 10% over this 5 year period—as occurred (roughly speaking) in the mid-1870s and again in the mid-1880s—the amount the farmer must repay measured in terms of goods is 10% higher than it was when he took out the loan. The prices of the farmer's own output are likely to have fallen by roughly this same amount over the period of the loan. The farmer must then sell 10% more wheat, corn, cotton, or whatever else he or she produces to be able to pay off the

loan, a considerable burden. In reality, agricultural prices fell by more than prices in general in the 1870s and 1880s, adding further to the plight of indebted farmers.

For these reasons, indebted farmers in the 1880s and 1890s were susceptible to propaganda about a “crime of 1873” perpetrated by eastern and foreign capitalists (“Wall Street”) to demonetize silver so as to raise the value of their loans at the expense of the public (“Main Street”). Others, who didn’t believe that a literal crime had been committed, also came to see the Coinage Act of 1873 as a “grave wrong” in terms of its economic consequences. William Jennings Bryan’s campaign to restore “free silver” at “16 to 1” was about reversing this wrong. His goal was to end the persistent deflation the U.S. had experienced over the prior 30 years and bring about inflation to ease the burdens of indebted farmers.

Other related ideas likely also factored in Bryan’s thinking. “Tight money” policies tend to depress output in the short run, as prices respond sluggishly to increased money demand. This occurs for reasons similar to those that produced the boom in output in response to the arrival of the viking gold in section 5. Bryan may have understood this at some level and wanted to reverse it.

The idea of restoring free coinage of silver at a price ratio of 16 to 1 was by many considered a thoroughly irresponsible proposal in 1896. At the time, the gold-silver price ratio was roughly 30 to 1. Opponents argued that Bryan’s policies would result in runaway inflation. A naive view was that the price level would roughly double, to reflect the low market price of silver. This, however, failed to take account of the fact that if the U.S. were to restore free coinage of silver at 16 to 1, the price of silver would have risen substantially.

It is not out of the question that the price of silver in terms of gold would have risen all the way to 16. After all, France had managed to stabilize the gold-silver price ratio in the vicinity of 15.5 from 1803 until the 1870s. Perhaps the U.S. could have done the same at the turn of the 20th century. In that case, the effect of adopting bimetallism in 1896 at 16 to 1 would have been modest. Friedman (1990b) argues, however, that a more likely outcome would have been a price ratio of about 24 in 1896 if Bryan had been elected (see also Friedman, 1992, ch. 4). In this case, Bryan’s election would have resulted in considerable inflation.

Curiously, despite Bryan losing the election of 1896, the deflation he sought to end did in fact end around that time. In the late 1890s, the price level reached bottom and gradually started to rise (see Figure 8). As a consequence, Bryan’s populist movement lost considerable steam and Bryan lost the presidential elections of 1900 and 1908 by wider margins than the election of 1896. The cause of the turnaround

in the price level is pretty obvious. The world production of gold skyrocketed in the 1890s. In the 1880s, about 50 tons of gold were mined annually in the world. This rose to roughly 130 tons in the period 1900-1914 (Eichengreen and McLean, 1994).

This huge increase in gold production was made possible by two roughly contemporaneous developments. First, vast gold fields were discovered in South Africa around this time. Second, a commercially viable process was developed to extract gold from the low-grade ore in these gold fields—the so-called cyanide process. So, it came to pass that William Jennings Bryan's chances of becoming president of the United States were undone by Scottish chemists and South African miners. Of course, the only reason why this could happen was that the workings of the monetary system of the United States at the time was subject to the vagaries of the global gold-mining industry.

Scholars who advocated bimetallism in the late 19th century (and 20th century) did so because they thought that bimetallism was likely to result in more stable prices than a monometallic system (such as a gold standard or silver standard). This represented a sharp break from earlier analysis of bimetallism. Prior to the 19th century, the primary rationale for bimetallism had been providing for an efficient medium of exchange (the big problem of small change discussed in section 8). But by the mid-19th century, this problem had been solved and price stability had become the paramount concern of monetary scholars. (Later scholars have sometimes scoffed at the idea that considerations of small change could drive the nature of the monetary system (Friedman, 1990a, p. 99-100).)

A bimetallic standard can result in a more stable price level than a monometallic standard because the second metal can add elasticity to the supply of money in cases where there is substantial mismatch between changes in money demand and changes in the supply of the first metal. Friedman (1990b) makes this case for the period 1875 to 1895. In particular, he argues that the U.S. would have been better served if the Coinage Act of 1873 had not omitted free coinage of silver. In that case, the U.S. would have resumed specie payments in 1879 on the basis of silver, as opposed to gold. This would have resulted in less world demand for gold over the subsequent decades and more world demand for silver. As a result the price level in gold standard countries would have fallen less and the price level in silver standard countries increased by less. Friedman argues that the price level in the United States would have been more stable over the period 1875 to 1914 in this case than it was in actuality. However, he stresses that this is a judgment about 1873, not 1896. By 1896, when William Jennings Bryan ran for president on a free-silver ticket, it was



too late. The damage of omitting silver in 1873 had already been done and could not be undone. As Friedman puts it: "Bryan was trying to close the barn door after the horse had been stolen."

## Notes

<sup>1</sup>Glassman and Redish (1985, p. 40) average two earlier estimates of the stock of money in Europe in 1492 from King (1696) and Del Mar (1877). The average is 15.1 million marcs *argent-le-roi*. They explain that 1 pound sterling is 0.4548 marcs and 1,718 troy grains of pure silver. One troy grain is equal to 0.0647989 grams. Marcs being *argent-le-roi* means that they are 23/24 pure silver. We therefore have  $15.1 \times 1,718 \times (1/0.4548) \times 0.06479 \times (23/24) = 3,541$ .

Velde and Weber (2000) estimate the value of the world stock of gold and silver in 1492 to be equal to the value of 6,900 metric tons of silver. They estimate the world stock of silver in 1492 to be 3,600 metric tons and the world stock of gold to be 297 metric tons. I use a price of gold in terms of silver of 11 to calculate the value of the stock of gold in terms of silver.

<sup>2</sup>The unite weighted 140.8 troy grains of 11/12 fine gold. Each troy grain is 0.0647989g. Each troy ounce is 480 grains. See Redish (2000, p. 27-28).

## References

- ALLEN, R. C. (2001): "The Great Divergence in Wages and Prices from the Middle Ages to the First World War," *Explorations in Economic History*, 38, 411–447.
- BARNETT, P. S. (1964): "The Crime of 1873 Re-examined," *Agricultural History*, 38.
- BLINDER, A. S., E. R. CANETTI, D. E. LEBOW, AND J. B. RUDD (1998): *Asking About Prices*, New York, NY: Russell Sage Foundation.
- BODIN, J. (1568): *La Response de Maistre Jean Bodin, advocat en la Cour, au Paradoxe de Monsieur de Malestroit touchant l'encherissement de totes hoses et le moyen d'y remedier*, Paris.
- (1606/1962): *The Six Bookes of a Commonweale*, Cambridge, MA: Harvard University Press.
- BUDISH, E. (2018): "The Economic Limits of Bitcoin and the Blockchain," Working Paper, University of Chicago.
- CIPOLLA, C. M. (1956): *Money, Prices, and Civilization in the Mediterranean World: Fifth to Seventeenth Century*, Princeton, NJ: Princeton University Press.
- (1976/1993): *Before the Industrial Revolution: European Society and Economy 1000-1700 (Third Edition)*, New York, NY: W.W. Norton & Co.
- DAVID, P. A. AND P. SOLAR (1977): "A Bicentenary Contribution to the History of the Cost of Living in America," 2, 1–80.
- DE MALESTROIT, S. (1566): *Les Paradoxes du Seigneur de Malestroit, conseiller du Roi et Maistre ordinaire de ses comptes, sur le fait des Monnoyes presented a Sa Majeste' au mois de Mars MDLXVI*, Paris.
- DEL MAR, A. (1877): "Minute on the World's Production of Gold and Silver," in *Re-*

- port of the U.S. Monetary Commission, Washington, DC: U.S. Government Printing Office.
- DOUGLAS, P. H. (1930): *Real Wages in the United States, 1890-1926*, Boston, MA: Houghton.
- EICHENGREEN, B. AND I. W. MCLEAN (1994): "The Supply of Gold under the Pre-1914 Gold Standard," *Economic History Review*, 47, 288–309.
- FEAVEARYEAR, A. (1963): *The Pound Sterling*, Oxford, UK: Oxford University Press.
- FRIEDMAN, M. (1990a): "Bimetallism Revisited," *Journal of Economic Perspectives*, 4, 85–104.
- (1990b): "The Crime of 1873," *Journal of Political Economy*, 98, 1159–1194.
- (1992): *Monetary Mischief*, New York, NY: Harcourt Brace & Company.
- FURNESS, H. W. (1910): *The Island of Stone Money: Uap and The Carolines*, Philadelphia, PA: J.B. Lippincott Co.
- GLASSMAN, D. AND A. REDISH (1985): "New Estimates of the Money Stock in France, 1493-1680," *Journal of Economic History*, 45, 31–46.
- HAMILTON, E. J. (1934): *American Treasure and the Price Revolution in Spain, 1501-1650*, Cambridge, MA: Harvard University Press.
- KING, G. (1696): *King, Natural and Political Observations and Conclusions upon the State and Condition of England*, London, UK.
- NAKAMURA, E. AND J. STEINSSON (2008): "Five Facts about Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415–1464.
- (2013): "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics*, 5, 133–163.
- OFFICER, L. H. (Undated-a): "What Was the Consumer Price Index Then? A Data Study," MeasuringWorth, <http://www.measuringworth.com/usdpi/>.
- (Undated-b): "What Was the Price of Gold Then? A Data Study," MeasuringWorth, <http://www.measuringworth.com/usdpi/>.
- OFFICER, L. H. AND S. H. WILLIAMSON (2021): "The Price of Gold, 1257-Present," MeasuringWorth, <http://www.measuringworth.com/gold/>.
- PIAZZESI, M. AND M. SCHNEIDER (2018): "Payments, Credit, and Asset Prices," Working Paper, Stanford University.
- REDISH, A. (1990): "The Evolution of the Gold Standard in England," *Journal of Economic History*, 50, 789–805.
- (2000): *Bimetallism*, Cambridge, UK: Cambridge University Press.
- REES, A. (1961): *Real Wages in Manufacturing, 1890-1914*, Princeton, NJ: Princeton University Press.

- SARGENT, T. J. AND F. R. VELDE (2002): *The Big Problem of Small Change*, Princeton, NJ: Princeton University Press.
- SMITH, A. (1776/2000): *An Inquiry into the Nature and Causes of the Wealth of Nations*, New York, NY: The Modern Library.
- SWEENEY, J. AND R. J. SWEENEY (1977): "Monetary Theory and the Great Capitol Hill Baby Sitting Co-op Crisis," *Journal of Money, Credit and Banking*, 9, 86–89.
- TEPASKE, J. J. (2010): *A New World of Gold and Silver*, Leiden, The Netherlands: Koninklijke Brill NV.
- VELDE, F. R. AND W. E. WEBER (2000): "A Model of Bimetallism," *Journal of Political Economy*, 108, 1210–1234.
- WALTER, J. R. (1989): "Monetary Aggregates: A User's Guide," Federal Reserve Bank of Richmond, Economic Review, Jan/Feb 1989.
- WILLIAMSON, S. H. (2021): "The Annual Consumer Price Index for the United States, 1774-Present," MeasuringWorth, <http://www.measuringworth.com/usdpi/>.