The Neutrality of Money. The term ‘neutrality of money’ has had numerous meanings over the years. Patinkin (1987) traces the entire history of its use. Currently, the term is used to in two specific ways. The first refers to the division of a static economic model into a real part, in which the quantity of output is determined, and a nominal part, in which nominal prices and wages are determined given the real quantities from the first part. In the second instance, the neutrality of money has come to mean that all anticipated monetary and fiscal policy — all aggregate demand policy — has no effect on output and employment in either the short or the long-run. The remainder of this essay will present a discussion of the theoretical and empirical issues surrounding these two meanings.

Monetary neutrality in a static macroeconomic model is synonymous with the term ‘classical dichotomy’. Standard models, such as Sargent (1986, Chapter 1) exhibit this property in which changes in the quantity of money generate proportional changes in all nominal variables in the economy, leaving real quantities unchanged. In other words, the model is homogeneous of degree zero in all nominal prices, nominal wages, and nominal money.

The following is a very simple version of such a model:

\[ y = f(n) \quad \text{Production function} \]  
\[ n^d = f'\left(\frac{W}{P}\right) \quad \text{Labor Demand} \]  
\[ n^s = g\left(\frac{W}{P}\right) \quad \text{Labor Supply} \]  
\[ y = c + i + \bar{g} \quad \text{National Income Accounting Identity} \]  
\[ c = c(y) \quad \text{Consumption function} \]  
\[ i = i(r) \quad \text{Investment function} \]  
\[ \frac{M^d}{P} = L(y, r) \quad \text{Money demand function} \]  
\[ \frac{M^s}{P} = \bar{M} \quad \text{Money supply function} , \]

where \( y \) is real output, \( n \) is the level of employment, \( W \) and \( P \) are nominal wages and prices, \( c \) and \( i \) are real consumption and investment, \( g \) is government expenditure on goods and services, \( r \) is the real interest rate, and \( M \) is the nominal money stock. The superscripts
and $s$ refer to demand and supply, the overbar signifies that a variable is determined exogenously, and $f^{-1}$ is the inverse of the first derivative of the production function.

If we assume that the labor market clears, and so $n^s = n^d$, equations (1), (2) and (3) determine the level of employment, output, and the real wage, without reference to either the level of government expenditure or the stock of nominal money. The real and nominal parts of this system are separate. Furthermore, changes in money have no impact on any real variables. Nominal prices and nominal wages will move so that the real wage and the level of real money balances are left unchanged.

Thus far, the discussion has been carried on in terms of a timeless, static model of the macroeconomy. It is now useful to introduce time, and distinguish between short-run and long-run neutrality. Let us think of the short run as several years and the long run as a limiting concept. Until recently, there was consensus that money was neutral in the long-run. That is to say the impact of all monetary changes on output died out, and so changes in the quantity of money result solely in changes in prices. Current developments in growth theory, primarily Romer’s (1986) endogenous growth model, imply that short-run output declines, i.e. recessions, may cause decreases in (human) capital accumulation, which has permanent consequences. I should also note that there is little consensus on the issue of whether money is ‘superneutrality.’ The term superneutrality refers to the case in which real quantities are unaffected by changes in the growth rate of money. There are clearly reasons to believe that changes in money growth rates can have real effects, as they create inflation, which is a tax on real balances, and they force agents to use more resource in changing their prices.

The remainder of this article is devoted to the theoretical and empirical case for and against short-run neutrality. The work of Lucas (1973), Barro (1978) and new other new classical macroeconomists has extended the concept of the neutrality of money to mean that no anticipated changes in money, or any other aggregate-demand policy variable, does not change the level of real output in the short-run. Their argument combines market clearing with rational expectations to yield the result that it is solely unanticipated policy, such as unexpected monetary shocks, that can affect output. Furthermore, by assuming that the markets for goods and labor are competitive, and clear each period, these researchers
conclude that these shocks should be short lived.

It is useful to phrase the issue in the language of macroeconomics textbooks. The proposition that money is neutral in the long-run is the statement that shifts in aggregate demand do not effect the location of the long-run aggregate supply curve. Put another way, current changes in the quantity of money leave the natural rate of unemployment unaffected. Thus, the statement that money is neutral in the short run means that, following a disturbance, the short-run aggregate supply curve shifts quickly to its long-run position.

The empirical evidence for short-run neutrality is extremely weak. Cecchetti (1986) derives a test under very general conditions that requires only that the econometrician specify the length of the maximum lag with which unanticipated money should influence output. The finding that money growth lagged is correlated with output movements contradicts the new-classical proposition that aggregate-demand policy is neutral in the short run. It is worth noting that the evidence of long-horizon effects of money on output is prima facie evidence in favor of the long-run nonneutralities emphasized in endogenous growth theory.

The new Keynesian economics, surveyed in Ball, Mankiw, and Romer (1989), provides the main alternative to the new classical proposition of short-run neutrality. While continuing to maintain the assumption that expectations are rational, new Keynesians examine the implications of a monopolistically competitive structure in which changing prices is costly, and prices change at different times. Nominal rigidities are important for two reasons. First, the fact that changing prices is costly creates an ‘aggregate-demand externality’; and second, the assumption of staggered price setting yields a mechanism for monetary shocks to be propagated over time. The remainder of this essay discusses each of these nonneutralities.

To understand the source of the aggregate-demand externality, assume that each firm’s demand depends on its relative price and the level of aggregate real balances, which is a measure of aggregate demand. This can be written as

$$y_i^D = \left( \frac{P_i}{P} \right)^{-\epsilon} \left( \frac{M}{P} \right), \quad (9)$$

where $P_i$ is the firm’s price level, $P$ is the aggregate price level, $M$ is the stock of nominal money. Now consider an experiment in which $M$ drops, and $\epsilon$ is the relative price elasticity.
of demand. If the firm’s price is set at the profit maximizing level prior to the monetary contraction, then, as pointed out by Akerlof and Yellen (1985), according to the envelope theorem, the loss from not adjusting $P_i$ is second order. This non-adjustment at the firm level implies rigidity of the aggregate price level, and generates a first-order fall in real balances, reducing the aggregate demand for goods. While this externality generates first-order output fluctuations, Ball and Romer (1990) show that the welfare loss is also second-order. The aggregate demand externality clearly implies monetary nonneutrality, but it may not be important.

The second source of nonneutrality in new Keynesian models emphasizes the temporal propagation of monetary shocks. The issue here is to understand how a purely nominal disturbance could have long-lived real effects that resemble business cycle fluctuations. How is it that following a monetary contraction a recession can last for several years? Even if prices are changed frequently, asynchronous timing will cause shocks to affect output for a long time.

To understand how this works, begin by assuming the opposite. Assume that firms’ pricing decisions are perfectly synchronized. As is discussed in Ball and Cecchetti (1988) this is the only possibility under perfect competition, and so the results are analogous to those of the new classical macroeconomics. If all firms change prices on the first day of every month, then adjustment to any monetary shock will take at most one month. While the shock can change real balances during the month, every firm will adjust at the beginning of the next month, knowing that all other firms are doing the same. This price adjustment would eliminate the real balance effect of the shock. Lags for data collection in the real world may cause a one-month delay in this process. Nevertheless, the economy jumps quickly back to the optimal level and money is neutral in the short-run.

Now consider the alternative in which firms adjust every month, but half change prices on the first of the month, while the other half change on the fifteenth. The reasons for this could be that firms receive idiosyncratic shocks at different times, as suggested by Ball and Romer (1989), or that firms stagger decisions in order to collect information, as described in Ball and Cecchetti (1988). Now consider a monetary shock that comes on the 10th of the month. Clearly this shock was unanticipated by the group that adjusted on the first of the
month, and so their prices will differ from those they would have set had they known the shock was coming. By contrast, the group adjusting on the fifteenth has the opportunity to take the shock into account. The assumption of imperfect competition implies that there is some substitutability between the products of the firms adjusting prices at different times, which is why the relative price appears in the demand function (9). That firms care about relative prices means that the firms adjusting on the fifteenth do not want their prices to move too far from the prices of the firms that adjust on the first. Thus, the expectation error built into the first group’s prices finds its way into the second group’s prices as well. Similarly, when the first group’s turn to change prices again arrives on the first of the next month following the shock, they will take into account the prices of the firms that last set their prices, which now include the expectation errors of their original price set before to the shock. In this way the impact of the shock dies out only asymptotically, with the effect peaking after all firms have adjusted prices once. The result is that nominal shocks have only short-run real effects.

Stephen G. Cecchetti
Bibliography


