FOUR PRINCIPLES FOR A BASE MONEY REGIME

William J. Luther

Most of the money used throughout the world is created inside the banking system. In the United States, for example, the M2 money stock less the currency component is nearly $13.5 trillion. The monetary base, in contrast, is only around $3.1 trillion. The difference is even greater in other countries, where financial institutions are not encouraged to stockpile so many reserves and the currency circulates to a much lesser extent beyond the borders of the issuing country or currency area. It might seem reasonable, then,

---

1The M2 money stock consists of currency outside the U.S. Treasury, Federal Reserve Banks, and the vaults of depository institutions; traveler’s checks of non-bank issuers; demand deposits; and other checkable deposits; savings deposits (including money market deposit accounts); time deposits in amounts of less than $100,000; and balances in retail money market mutual funds.

2The monetary base includes currency in circulation outside Federal Reserve Banks and the U.S. Treasury as well as deposits held by depository institutions at Federal Reserve Banks.

3Selgin (2018a) explains how the Federal Reserve encourages financial institutions to stockpile reserves. Rogoff (2016) surveys the literature considering the extent to which various currencies circulate abroad.
to focus on bank-issued money and pay little attention to base money. But that would be a mistake.

The claims issued by banks, primarily demand deposit account balances, are redeemable for base money. As such, these claims derive their value from the underlying base money asset, perhaps with some discounting for bank-specific risks. A bank that issues too few claims leaves profits on the table.\(^4\) Since creating and lending additional claims generates sufficient income to warrant the additional risk of illiquidity, such a bank would be well-served by decreasing its reserves to deposits ratio. A bank that issues too many claims, in contrast, takes on more risk of illiquidity than is warranted by the income those claims generate. Such a bank would be well-served by increasing its reserves-to-deposits ratio, since it risks suffering reserve drain and losing market share to rival banks. Base money is high-powered money. It serves as the reserves of the banking system. An optimizing bank will only expand its loans and deposits as the supply of base money expands or demand for base money contracts; it will only contract its loans and deposits as the supply of base money contracts or demand for base money expands.\(^5\)

The contractual obligation to redeem its claims for base money constrains a bank. Hence, the base money regime is of the utmost importance. The growth of base money—and, in particular, reserves—determines the course of financial activity. Bank-issued money is the ship. Base money is the rudder.

In what follows, I offer four principles for a base money regime. In brief, I maintain that an ideal base money is (1) stable, (2) demand elastic, (3) global, and (4) incentive compatible. I argue that these principles are necessary for evaluating alternative base money regimes, like a government-issued fiat money, commodity money, or more recent cryptocurrencies. I conclude that, since most real-world monies fall short of the ideal on one margin or another, one must consider the tradeoffs when ranking the available alternatives.

\(^4\)In the United States, for example, bank-issued deposit account balances are redeemable for U.S. dollar currency or U.S. dollar reserves held at the Federal Reserve. Since the Federal Reserve is committed to maintaining parity between reserves and currency, it is reasonable to think of U.S. dollar currency and reserves as the same asset.

\(^5\)For a more complete presentation of this view, see Selgin (1988); Selgin and White (1994); and White (1999: 53–67).
Long-Run Stability

How should the supply of money be governed? I set aside, for now, consideration of the precise mechanism or the features of a desirable mechanism. Instead, I focus on the outcomes of a desirable mechanism. I also set aside the question of how the money supply should behave over relatively short periods of time. For now, I focus exclusively on how the money supply should behave over the long run. To that end, I take the consensus view—that the absolute rate of money growth should be low—and consider the range of disagreement in the literature.

In the long run, inflation is a monetary phenomenon. Hence, asking how the money supply should be governed over the long run is akin to asking what the optimal rate of inflation is over the long run, where “optimal” means maximizing genuine productivity.6

There are essentially three views in the literature: (1) zero inflation, (2) slightly negative inflation, and (3) slightly positive inflation. The first view maintains that the productivity-maximizing, or optimal, rate of inflation is zero. Advocates of such a position, including Meltzer (1999), typically maintain that the general price level is a numeraire and, hence, any costs incurred to change the numeraire are unwarranted. Some also note that, since nominal capital gains are subject to tax, inflation results in costly distortions in saving and investment (Feldstein 1999). Still others stress the salience of zero inflation, which might reduce uncertainty (and, hence, boost productivity) about price level drift (Gavin and Stockman 1988).

The second view maintains that inflation should be slightly negative. This view, credited to Friedman (1969), is typically justified on the grounds that individuals will hold insufficient cash balances when the real rate of return on currency is lower than the real rate of return on bonds of similar risk and duration. If individuals do not hold enough cash, they will have to incur the costs of managing smaller cash balances. Moreover, some transactions (and, hence, the mutual gains from those transactions) will go unrealized. Paying interest on cash is technically possible (e.g., banknote lotteries), but very costly.

6To say that the optimal rate of inflation is the rate that makes society the most productive is not to say it is the rate that causes society to produce the most. Monetary mischief might fool people into overproducing. But overproduction is unproductive, as it wastes valuable resources. Rather, the optimal inflation rate promotes genuinely productive ventures.
Far better, proponents of the second view maintain, to generate a mild, expected deflation so that currency yields a positive real rate of return comparable to similar financial assets. Then, everyone will hold enough cash to make the desired transactions.

The third view maintains that some low, but positive rate of inflation is desirable. This view, as put forward by Akerlof, Dickens, and Perry (1996), is usually justified on the grounds of lubricating labor markets—that is, enabling employees to accept a lower real wage without enduring the psychic costs of seeing the amount on their paychecks decrease and enabling employers to offer lower real wages without reducing morale and, hence, labor productivity. Others more or less accept the first view but worry that conventional measures overestimate inflation (Bernanke and Mishkin 1997). In other words, achieving zero percent actual inflation requires a slightly positive rate of measured inflation. More recent studies yield a positive optimal rate of inflation on the grounds of avoiding the zero lower bound (Carreras et al. 2016) or compensating for financial frictions (Finocchiaro et al. 2018).

Note that, in all three views, there is some cost to be minimized. In the first view (and the third view, when held on the grounds of mismeasurement), it is the cost of changing prices. In the second view, it is the costs of managing smaller cash balances and occasionally missing out on exchange opportunities. In the third view, it is transactions and/or psychic costs. A more general view would take all of these costs (and potentially others) into account (e.g., Aiyagari 1990). But the underlying logic is the same. Incurring costs means using scarce resources. If the costs can be reduced on net, resources are freed up to produce other valuable goods and services, thereby increasing overall productivity.

While the preceding logic provides a useful starting point, it would be more useful still to be precise about the range of optimal inflation rates advanced by those working in the field and the extent to which they accept those estimates. A recent survey of the literature by Diercks (2019) allows for both. He finds that most papers—and the most-cited ones—recommend an inflation rate around 0 percent.

---

7See also Boskin et al. (1996).
8Schmitt-Grohe and Uribe (2012) suggest this view might be misguided.
The average inflation rate across all 159 papers surveyed is 0.03 percent. Many papers find that a slightly negative rate of inflation is desirable. Some papers, most of which have been published in the last 10 years, find that a slightly positive rate of inflation is ideal. Only 10 papers (6.29 percent) recommend an inflation rate in excess of 4 percent. Only one paper (0.63 percent) recommends an inflation rate below −4 percent.

Taken together, the consensus view holds that the absolute rate of money growth should be low over the long run. The optimal inflation rate is almost certainly between −4 percent and 4 percent, and probably much closer to zero. And the optimal rate of base money growth must be consistent with that outcome.

Short-Run Variability

Having considered how the supply of base money should be governed over the long run, I now turn to the short run. I maintain that the ideal base money is demand elastic—that is, the supply of base money expands and contracts to offset changes in the demand to hold money.\textsuperscript{10}

Consider, first, the case for a demand-elastic base money when prices are perfectly flexible. The equation of exchange identifies that \( MV = PY \), where \( M \) is the money supply, \( V \) is the income velocity of money, \( P \) is the price level, and \( Y \) is real gross domestic product. For simplicity, we assume the demand to hold money is \( M_D = kY \), where \( k \) is desired fluidity and, hence, \( V = \frac{1}{k} \). When prices are perfectly flexible, \( Y = f(A, K, L) \), where \( A \) is the level of technology, \( K \) is the physical capital stock, and \( L \) is the human capital stock. In other words, real output depends on the level of technology and the factors of production.

An increase in the demand to hold money (i.e., desired fluidity \( k \)) is a decrease in velocity, \( V \). If prices are perfectly flexible, they will adjust from \( P_1 \) to \( P_2 < P_1 \). Since there has been no change in real fundamentals, \( Y \) is unchanged and the relative scarcity is conveyed equally well by the price vectors \( P_1 \) and \( P_2 \). In other words, there is no real benefit of allowing the price level to adjust. Hence, when prices are perfectly flexible, it is preferable to maintain equilibrium

\textsuperscript{10}Selgin (2018b) offers a more complete argument for this position.
by adjusting the money supply so long as $C_M < C_P$, where $C_M$ is the cost of adjusting the money supply and $C_P$ is the cost of allowing the price level to adjust (Luther and Salter 2012).

Of course, few economists maintain that prices are perfectly flexible. Instead, they hold that exchange (Mankiw 1985) or epistemic frictions (Lucas 1972; Mankiw and Reis 2002) cause prices to adjust with a considerable lag. Let $Y_N$ denote the natural rate of output, where $Y_N = f(A, K, L)$, $Y = Y_N + a[P - E(P)]$, $a$ is a sensitivity parameter, and $E(P)$ is the price level expected prior to the nominal disturbance. In other words, changes in the demand to hold money will result in short-run over- or underproduction until prices have sufficient time to adjust.

When prices are sticky, the case for a demand-elastic base money is even stronger. Recall that, in the case of a money demand shock when prices are flexible, the real economy performed just as well regardless of whether the price level was permitted to adjust. When prices are sticky, in contrast, the real economy performs worse if the price level adjusts in response to a money demand shock. Hence, it is preferable to maintain equilibrium by adjusting the money supply so long as $C_M < C_P + C_Y$, where $C_Y$ denotes the cost of temporarily over- or underproducing.

Should the supply of money adjust in response to real supply shocks? Recall that, in response to a money demand shock, the initial price level, $P_1$, conveys information about relative scarcity just as well as the alternative price level, $P_2$. Following a real shock, however, relative prices must adjust to reflect the new real underlying fundamentals.

Real shocks tend to be concentrated on a particular subset of markets. Allowing the price level to adjust would require price adjustments only in those markets affected. Maintaining the initial price level by adjusting the money supply, in contrast, would require changing all prices—a costlier proposition. Hence, $C_{P(R)} < C_{P(M)}$, where $C_{P(R)}$ denotes the cost of price adjustments necessitated by the real shock and $C_{P(M)}$ denotes the cost of price adjustments necessitated by the money supply adjustment to preserve the initial price level.

Moreover, since real shocks tend to be concentrated on a particular subset of markets, information concerning the need to adjust prices tends to be concentrated on those markets; and the relevant actors in those markets have an incentive to change their bidding or
asking prices accordingly. So prices will tend to be less sticky in those markets where adjustments need to be made following a real shock. And, by requiring price adjustments in other markets, adjusting the money supply is likely to induce over- or underproduction in those other markets, where market participants are more likely to be fooled when encountering the signal extraction problem. More formally, I maintain that $C_P(R) + C_Y(R) < C_M + C_P(M) + C_Y(M)$, where $C_Y(R)$ is the cost of over- or underproducing following the real shock and $C_Y(M)$ is the cost of over- or underproducing following the money supply adjustment.

The Ideal Base Money Is Global

Money is an institutional technology that promotes trade. It enables users to overcome the difficulties of barter (Menger 1892) when trust is insufficient (Kiyotaki and Moore 2002) at a relatively low information cost (Brunner and Meltzer 1971; Alchian 1977). More generally, it allows a society to achieve a more desirable allocation than could be achieved without money (Hahn 1987; Lagos and Wright 2008). When considering the desirability of alternative base monies, then, one should not forget the extent to which those monies facilitate exchange.

If facilitating exchange were the only issue, the ideal base money would circulate as widely as possible. A common currency reduces information costs by making cross-border comparisons less onerous. It reduces transaction costs by eliminating the need to convert prices or exchange one money for another before transacting. And it reduces the riskiness of cross-border production plans since, under a common currency, such plans are not subject to exchange rate risk. As a result, a common currency promotes exchange both directly, by lowering the costs of exchange, and indirectly, by increasing productivity. Hence, barring interstellar considerations (Krugman 2010), the ideal base money as evaluated solely on the grounds of facilitating exchange would be global.

The standard view maintains that facilitating exchange is not the only issue, however, since expanding the domain of a base money also expands the domain of monetary policy—or, more generally, the mechanism to govern the supply of base money. A money circulating over the most expansive region possible cannot have a money-supply mechanism tailored to the various constituent regions. If the
marginal benefit of expanding the circulation—that is, the gains from additional exchanges made possible by a common currency—is diminishing while the marginal cost—that is, the losses from the less-well-tailored common money-supply mechanism—is increasing, the optimal currency area might be smaller than the maximum currency area (Mundell 1961).

In considering the size of the optimal currency area, I take it for granted that the money-supply mechanism will be conducted as well as is possible over the currency area. The only question, then, is whether a more narrowly tailored money-supply mechanism would improve matters. I argue that it would not.

Contrary to chalkboard abstractions, nominal shocks do not hit all people in a given region equally. A general increase in the demand to hold money, for example, results from the specific increases of specific people in specific places (perhaps mitigated by the specific decreases of others), which almost certainly vary by degree. Each person demanding more money will have a specific willingness to pay for that money. The financial system will tend to allocate the available funds to the highest bidders (Luther and Salter 2012). Hence, a general increase in the supply of money to offset a general increase in money demand will tend to channel funds to the specific people demanding more money (Harwick 2017). A common currency area can have but one money-supply mechanism. However, that money-supply mechanism gets tailored to the needs of the constituent regions through the financial system.

What about regional real shocks? Price-level differences between regions that are consistent with real underlying fundamentals are relative price differences. They convey important information about relative scarcity and, hence, should not be seen as a problem. The cost of living is lower in Montgomery, Alabama (population: 198,218) than in New York City (population: 8,398,748). But you would have to live in Montgomery to get the discount. The lack of folks moving from New York to Alabama suggests few think the discount is large enough to warrant their expected decline in quality of life. No economist is puzzled by the difference in price between ground beef and steak. Why, then, do some worry about the difference in the cost of living across the constituent regions of a common currency area?

Likewise, changes in the real underlying fundamentals should be reflected by changes in regional price levels. If there is a drought in Nebraska, for example, farmers in Nebraska will produce less corn.
Ranchers there will pay more for seed corn, both because domestic corn is more scarce and there are additional transport costs for imported substitutes. Prices will rise in general (i.e., in Nebraska and elsewhere). But the effect will be especially pronounced in Nebraska, since it is relatively less productive following the drought. And there is nothing that changing the supply of money can do to improve matters since Nebraskans really do have less purchasing power. They are just not as productive as they would have been without the drought.

The benefits from exchange increase with the circulation of a base money. A common currency area prevents a regionally specific money-supply mechanism. However, the financial system routes funds to those demanding them and remaining differences in regional price levels are relative-price differences. Taken together, this implies that the optimal currency area is the maximum currency area. In other words, the ideal base money is global.

An Incentive-Compatible Base Money Requires a Credible Commitment

A regime is incentive compatible if and only if it is in the interest of the relevant actors to carry out the inner workings of the regime as described. If a regime is not incentive compatible, there is little reason to expect that the regime will behave as described. It is also less likely to persist. Hence, if a regime is not incentive compatible, it should be evaluated both as it is described and as it will likely behave, weighting both according to their likelihood in each period and discounting future values accordingly.

Adhering to the first three principles identified herein is only sufficient to capture a portion of the available benefits. Capturing all of the benefits requires an observable credible commitment to those principles. And a commitment is only credible if it is incentive compatible.

Consider the benefits that come from long-run stability. Credibly committing to such a supply mechanism anchors expectations, enables long-term contracting, and, hence, promotes economic growth. If the commitment is not credible, long-term contracting is riskier. Some will incur costs to mitigate that risk, perhaps by acquiring additional information or switching from projects that require long-term contracts to more costly short-term alternatives. These costs divert resources from other productive uses.
The benefits of a global monetary regime, similarly depend on the extent to which it is incentive compatible. Recall that one of the benefits of a global monetary regime is the reduction in exchange rate risk. However, if such a regime is not incentive compatible, those considering cross-border production plans must consider the risk that exchange rate risk will become an issue over the relevant time horizon, perhaps because the regime is abandoned or otherwise devolves into a riskier alternative. Hence, some of the potential benefits of a global regime will go unrealized if that regime is not incentive compatible.

An ideal base money regime, therefore, is incentive compatible. It is not enough for the relevant actors to merely do the right thing. There must be a credible commitment to do the right thing. Otherwise, some of the available benefits will go unrealized.

Conclusion

There is a sense in which the status quo is (axiomatically) optimal. If something better were possible, given the prevailing constraints, it would have been achieved. Those suggesting otherwise merely overlook some relevant cost (Leeson forthcoming). But such a view does not preclude one from attempting to improve the world. One need only make decisions today that sufficiently relax the future constraints of the relevant decisionmakers. Optimizers optimize; and, facing the less restrictive constraints, they will make decisions that yield more social welfare than before.

One such constraint is the state of ideas. If the relevant decision-makers do not know that an alternative is superior, they cannot be expected to choose it. Indeed, the first step toward institutional change is to demonstrate that one alternative is superior to another. And the first step toward demonstrating that one alternative is superior to another is establishing a clear benchmark against which the available options can be judged. I have offered four principles for a base money regime.

In the long run, it should be stable, yielding an inflation rate of around 0 percent. In the short run, it should expand and contract to accommodate changes in money demand. It should be global, thereby facilitating trade to the maximum extent possible. And, finally, it should be incentive compatible, so that the expectations-based benefits of the three preceding principles can be realized.
The available real-world alternatives fall short of the ideal on one or more margins. For example, the historical experience with government-issued fiat monies suggests they are far from stable. They have yielded significantly higher rates of inflation than commodity money regimes (Rolnick and Weber 1997). Commodity monies, however, have historically been slow to adjust in the short run (Bordo 1984: 217, n. 36). Cryptocurrencies like bitcoin have the potential to facilitate exchange well beyond national borders without the international coordination required by traditional monetary unions. However, their supplies are often even more rigid than traditional commodity monies (Cachanosky 2019; Caton forthcoming). Even though the available real-world alternatives fall short of the ideal, the four principles outlined herein provide a starting point for evaluating the alternatives. We are in a world of second (third or fourth) best and must consider the relevant tradeoffs. But, without a clear ideal in mind, such comparisons would be impossible.

References


