Economic Review

FEDERAL RESERVE BANK OF DALLAS
FOURTH QUARTER 1995

In the aftermath of recent economic events in Argentina and Mexico, economists are asking whether currency boards are an appropriate tool for maintaining monetary stability. For some, the issue amounts to a question of rules versus discretion.

Economic Review
Federal Reserve Bank of Dallas

The *Economic Review* is published by the Federal Reserve Bank of Dallas. The views expressed are those of the authors and do not necessarily reflect the positions of the Federal Reserve Bank of Dallas or the Federal Reserve System.

Subscriptions are available free of charge. Please send requests for single-copy and multiple-copy subscriptions, back issues, and address changes to the Public Affairs Department, Federal Reserve Bank of Dallas, P.O. Box 655906, Dallas, TX 75265-5906, (214) 922-5257.

Contents

This article by John Duca discusses how shifts in technology, transactions, and asset preferences can weaken the relationships between monetary aggregates, the opportunity cost of money, and nominal output. Observed shifts in these general relationships are shown to be consistent with plausible changes in technology and preferences. Evidence indicates that technological advances have reduced the costs of shifting across assets and have lowered the precautionary need to hold monetary assets as a means of conducting transactions. Aside from technological changes, demographic and employment shifts have boosted the role of households in directing investments earmarked for funding their retirement and may have thereby increased their tolerance for investment risk. In turn, these factors may have induced households to shift their portfolios from monetary assets toward riskier assets with higher expected long-run yields.

This article discusses currency boards in light of the recent economic experiences of Mexico and Argentina. Carlos Zarazaga argues that currency boards do not solve the important time inconsistency problem pointed out in the rules-versus-discretion literature. Because of this failure, even the quasi-currency board established by law (the so-called convertibility law) did not protect Argentina from one of its most severe financial crises in modern times.

In addition, there is the normative issue of whether an ironclad rule such as a currency board rule is superior to a noncontingent one. Zarazaga argues that is not the case, except perhaps when the distinction between these two kinds of rules has become blurred in countries with poor reputations for following policy commitments. In such circumstances, ironclad rules theoretically might be desirable, although this conjecture has yet to be proved formally and verified empirically. Zarazaga argues that the study of the recent economic experiences of Mexico and Argentina could be useful for addressing both issues.

This article examines the effects and desirability of paying interest on required reserves. Scott Freeman and Joseph Haslag demonstrate that a policy of paying interest on reserves can make everyone better off, even if the interest must be financed by a tax on capital. An essential part of this policy is an open market operation that offsets any changes in the value of money.

Sources of Money Instability

John V. Duca
Research Officer
Federal Reserve Bank of Dallas

Recent studies generally conclude that the link between nominal output, interest rates, and conventional definitions of broad money has weakened or shifted. By reviewing the recent literature in the context of a microtheoretic model of money, this article attempts to shed light on why these relationships have changed.

During the early post–World War II era, the relationship between money and nominal output was stable, which encouraged many analysts to use money as an economic indicator. This reliance can be discussed using the equation of exchange:

$$(1) \qquad M \times V = P \times T,$$

where M = money, V = velocity [nominal gross domestic product (GDP)/M], P = the price level, and T = transactions (usually measured by inflation-adjusted GDP). Money holdings typically fall as the spread between a riskless short-term market interest rate and the average yield on monetary assets rises. As a result, the velocity of money rises as this spread, or opportunity cost of money, increases. If velocity is predictable, then money and its predicted velocity can be used to infer nominal GDP ($P \times Y$). Under these conditions, money is a useful indicator because data on GDP are available after a long lag, whereas information on money and interest rates is more timely.

Among active researchers, confidence in the M1 monetary aggregate (currency plus checking deposits) peaked with the publication of a money demand study by Goldfeld (1973). This study found that M1 reflected movements in nominal GDP and, to a smaller extent, changes in the three-month Treasury bill rate. These results implied that nominal GDP growth roughly equaled M1 growth, with a small adjustment for interest rates. Shortly after that study was published, M1 grew unusually slowly relative to nominal GDP, giving rise to Goldfeld's (1976) "case of the missing money."

In the early 1980s, the interest sensitivity of M1 jumped as financial innovations and deregulation created new deposits that combined savings and transactions features (see Hetzel and Mehra 1989) and helped firms avoid holding non-interest-bearing demand deposits (see Tinsley, Garrett, and Friar 1981). Partly as a result, attention shifted to M2, a less interest-sensitive and broader aggregate that was created in 1980 (see Simpson 1980). M2 was defined to include money market mutual funds (MMMFs) and overnight instruments, which became important in the late 1970s, and was redefined in 1982 to include money market deposit accounts (MMDAs). M2 had a reasonably stable relationship with interest rate spreads and nominal GDP during the 1980s (see Moore, Porter, and Small 1990 and Small and Porter 1989). However, this relationship broke down in the 1990s as M2 became more sensitive to long-term interest rates (see Duca 1995, Feinman and Porter 1992, and Koenig 1995a) and as households shifted toward bond and stock mutual funds (see Collins and Edwards 1994; Duca 1995, 1994a, 1994b; and Orphanides, Reid, and Small 1994) and toward direct holdings of Treasury securities (see Feinman and Porter 1992).

These breakdowns in the link between money and nominal output have spurred efforts to redefine money (for example, Collins and Edwards 1994, Duca 1995, and Hess and Morris 1995) or revise money models to account for changing behavior (for example, Koenig 1995a). Understanding why the money–income relationship has shifted is critical to finding new ways of deriving useful information from money and is the subject of this article.

The next section presents a simple theoretical model to illustrate three sources of change in the link between money and nominal output, followed by a section presenting evidence on shifts in this relationship. The subsequent three sections then review evidence on each potential source of money instability. The conclusion speculates on how likely changes in financial practices will affect money demand in coming years.

How the relationship of nominal GDP, interest rate spreads, and money can shift

In the early 1990s, households increasingly viewed nonmonetary assets as more attractive than M2 deposits for a given spread between expected yields on nonmonetary and monetary assets. As a result, conventional econometric models that use income and yield spreads to account for movements in money generally overpredicted M2 growth in that period, giving rise to "the case of missing M2." Theoretical models of money imply that the breakdown of such econometric models likely stems from their failure to control for other factors affecting money holdings. With respect to M2, these factors can be illustrated using a modified version of Milbourne's (1986) model of financial innovation and liquid assets.

Milbourne's framework is a modified Miller–Orr model (Miller and Orr 1966) in which households face uncertain net cash flows in a world with three financial assets: transactions accounts yielding a return of r_d, bank savings accounts yielding r_s, and bonds—which have virtually no credit risk—yielding r_b.[1] Changes in net cash flow are stochastic, with a mean of 0 and a variance of σ^2. Whenever transactions balances hit zero, funds are transferred at a fixed cost into transactions accounts from either savings accounts or from bonds. Milbourne assumes that $r_d < r_s < r_b$ (or more liquid assets yield lower pecuniary returns) and that the fixed cost of transferring funds from bonds into transactions accounts (β) is greater than that of shifting funds from savings to transactions accounts (α). Owing to the latter assumption, Milbourne's model implies that households will hold a portfolio of all three financial assets and that transactions deposits (D), small time deposits (S), and total M2 deposits ($M2 \equiv S + D$) equal

$$(2) \qquad D = (4/3)^{2/3}\sigma^{2/3}(\alpha/[r_b - r_d])^{1/3},$$

$$(3) \qquad S = (4/3)^{2/3}\sigma^{2/3}(\beta/[r_b - r_s])^{1/3}, \text{ and}$$

$$(4) \qquad M2 = (4/3)^{2/3}\sigma^{2/3}[(\alpha/[r_b - r_d])^{1/3} + (\beta/[r_b - r_s])^{1/3}],$$

respectively. Milbourne shows that with $r_b > r_s$, a rise in the cost of transferring funds from bonds to transactions accounts will, by making bonds more costly to hold, cause money balances to rise ($\partial\log(M2)/\partial\log(\beta) > 0$), which implies that a fall in β will lead to slower M2 growth.

This model can be modified in two ways to make it more relevant. First, note that, by definition, the standard deviation (volatility) of net cash flow (σ) rises with the average or expected level of transactions:

$$(5) \qquad \sigma = \gamma T,$$

where γ is the coefficient of variation. Equation 5 reflects that as the average levels of cash inflow and outflow rise with transactions in magnitude, so will the magnitude of the expected volatility (standard deviation) of net cash flow. To show a link with output (Y), assume that transactions are typically proportionate to output with some temporary deviations:

$$(6) \qquad T = Y(1 + \epsilon),$$

where ϵ has a mean of zero and a variance of var_ϵ.

The second major change is to treat bonds and equity as the alternative asset to D and S. Because bonds and equity carry price risk, replace r_b with a risk-adjusted expected return on bonds and equity (r_q):

$$(7) \qquad r_q = E(r_b) - bvar_{rb},$$

where the parameter b is the risk adjustment factor and var_{rb} is the variance of returns on stocks and bonds. The additive adjustment in equation 7 is consistent not only with the quadratic utility function used by Tobin (1958), which exhibits increasing risk aversion in wealth, but also with utility functions that are characterized by constant relative risk aversion, which is more consistent with empirical research (for example, Friend and Blume 1975) and with the common perception that risk aversion tends not to increase as wealth levels rise. As households become more risk averse, b rises in magnitude.

Substituting equations 5–7 into equation 4 yields

$$(8)\qquad M2 = (4/3)^{2/3}[\gamma Y + \gamma \epsilon Y]^{2/3} [(\alpha/[E(r_b) - bvar_{rb} - r_d])^{1/3} + (\beta/[E(r_b) - bvar_{rb} - r_s])^{1/3}].$$

Rearranging equation 8 produces

$$(9)\qquad Y^{2/3}/M2 = (4/3)^{-2/3}[\gamma(1 + \epsilon)]^{-2/3} \{(\alpha/[E(r_b) - bvar_{rb} - r_d])^{1/3} + (\beta/[E(r_b) - bvar_{rb} - r_s])^{1/3}\}^{-1},$$

which has several empirical implications about velocity (Y/M).

Most econometric models implicitly treat interest rate spreads as having a constant effect on velocity over time. Equation 9 implies that velocity may deviate from what these models predict for three reasons:

1. Deviations of output from transactions (ϵ) will introduce noise into the relationship between money and output.
2. Declines in the costs of transferring funds from nonmonetary assets to transactable assets (β) and from nonmonetary assets to nontransactions M2 accounts (α) will raise the velocity of M2.[2]
3. An increase in household tolerance for risk (that is, a decline in b) will lead to a rise in the velocity of money.[3]

In addition to these effects, another empirical implication arises.

4. Because econometric models do not have good time series measures of α, β, and b, a decline in any one of these parameters will (a) likely boost the estimated sensitivity of M2 to opportunity cost spreads as samples are extended into the 1990s and (b) affect the estimated sensitivity of M2 plus bond and/or equity mutual funds to a smaller extent because these expanded aggregates internalize most of the shifts between M2 and non-M2 that arise from changes in these parameters. (See the box entitled "Omitted Variable Bias.")

Omitted Variable Bias

According to the model presented in the article, the elasticity of transactions deposits with respect to their opportunity cost is constant (–1/3), as is the elasticity of small time deposits with respect to their opportunity costs.[1] However, the elasticity of M2 with respect to the opportunity cost of transactions deposits should be smaller in magnitude as the cost of transferring funds from small time deposits to transactions deposits (α) falls. In addition, a decline in the cost of transferring funds from nonmonetary assets to transactions deposits (β) will lead to a decline in the size of the elasticity of M2 with respect to the opportunity cost of small time deposits.[2]

The intuition for the first result is that as the cost of transferring funds between small time and transactions deposits falls, the transactions share of M2 falls. As a result, a given percentage increase in the opportunity cost of transactions accounts has a smaller impact on overall M2, even though it has the same percentage effect on transactions accounts. The same logic extends to the impact of a decline in β on the elasticity of M2 with respect to the opportunity cost of small time deposits. If both costs fall, the net effect on both elasticities is, a priori, ambiguous. Only if technological change is balanced, in the sense that the percentage changes in α and β are equal (that is, $\Delta\alpha/\alpha = \Delta\beta/\beta$), will the net change in each elasticity be zero.[3]

However, because econometric models do not, as of yet, have good time series measures of α, β, and b, a decline in one of these parameters will, over time, boost the estimated sensitivity of M2 to the spread between returns on nonmonetary assets and money.[4] Since most conventional models constrain the income elasticity of money to be constant when the models are estimated, the models will try to account for the negative impact of recent declines in transfer costs and risk aversion by boosting the size of the estimated negative coefficients on interest rate spreads. As a result of constraining the income coefficients to be constant over time, M2 will likely appear more sensitive to opportunity cost spreads in these models as samples are extended into the early 1990s, while this omitted variable bias is less of a problem for an aggregate that adds bond and equity funds to M2. In addition, because the yield curve was steep in the early 1990s, a wide spread between long-term interest rates and the average rate on M2 deposits (which tend to have short-term maturities) will be correlated with omitted data on declining transfer costs (and/or declining risk aversion). As a result, omitted variable bias will likely result in a rise in the observed sensitivity of M2 to long-term opportunity cost measures as samples are extended into the early 1990s.

1 The opportunity cost of transactions deposits in this model is the risk-adjusted spread between the expected return on securities (stocks and bonds) and the transactions deposit rate, while the opportunity cost of small time deposits is the risk-adjusted spread between the expected return on securities and the small time deposit rate.

2 Denoting the opportunity costs of transactions and small time accounts by c^d and c^s, respectively, the opportunity cost elasticities of M2 simplify to

$$\Delta\left[\partial M2/\partial c^d\right]\left[c^d/M2\right] = \left\{\frac{(1/9)\alpha^{1/3}\beta^{1/3}[c^s]^{1/3}[c^d]^{1/3}}{\{\alpha^{1/3}(c^s)^{1/3} + \beta^{1/3}(c^d)^{1/3}\}^2}\right\}\left\{\frac{-\Delta\alpha}{\alpha} + \frac{\Delta\beta}{\beta}\right\}, \text{ and}$$

$$\Delta\left[\partial M2/\partial c^s\right]\left[c^s/M2\right] = \left\{\frac{(1/9)\alpha^{1/3}\beta^{1/3}[c^s]^{1/3}[c^d]^{1/3}}{\{\alpha^{1/3}(c^s)^{1/3} + \beta^{1/3}(c^d)^{1/3}\}^2}\right\}\left\{\frac{\Delta\alpha}{\alpha} + \frac{\Delta\beta}{\beta}\right\}.$$

3 Balanced technological change does not affect the transactions and nontransactions shares of M2. Since the opportunity cost elasticities of transactions and nontransactions deposits are constants, the constant shares imply that the opportunity cost elasticities of M2 are also unchanged.

4 The same is true for estimates of the elasticity of transactions deposits with respect to the opportunity cost of transactions deposits and of the elasticity of small time deposits with respect to their opportunity cost.

The impact of financial churning, or volatility, falls under the first category. The second implication covers technological advances that lower α or β, such as declines in the costs of using mutual funds, the spread of automatic teller machines, improvements in services offered by mutual funds, and greater ease in purchasing Treasury securities. The third empirical implication reflects not only changes in risk aversion stemming from demographic and preference shifts, but also improvements in nonmonetary assets that make it easier for households to obtain a well-diversified portfolio and a greater awareness of alternative investments that makes households more willing to hold non-M2 assets. In practice, the fourth implication appears as omitted variable bias that leads to parameter instability in conventional money-demand functions.

Evidence on this instability is presented in the next section, partly as a means of showing why monetary economists are concerned about issues regarding financial technology, preferences, and volatility. Then, the subsequent sections review evidence on how technological changes, shifts in preferences and demographics, and volatility in financial transactions have affected the demand for money in ways not captured by conventional models.

Evidence suggestive of omitted variable bias

If substantial shifts in monetary technology and preferences have occurred, then we can observe two types of results from econometric models that do not contain good time series measures of transfer costs and risk adjustments to nonmonetary asset returns. First, expanding M2 to include some of these nonmonetary assets could yield an aggregate that would better predict inflation and nominal output in the 1990s and would have a more stable long-run relationship to those variables in money models than would the current definition of M2.[4] For this to occur, the advantage of internalizing time-varying substitution between M2 and such assets needs to outweigh any extra volatility that arises because the value of the added non-M2 assets fluctuates (that is, most of these non-M2 assets have price risk, unlike M2 components). If this condition holds, then a second type of result arises: the impact of asset returns in models of M2-like aggregates, inflation, and nominal output should vary less over time when a broader version of M2 replaces the current definition of M2 either as a dependent or independent variable.

Three types of evidence are consistent with these implications. First, econometric studies have found that adding bond and/or stock mutual fund assets to M2 yields an aggregate that has outperformed M2 in predicting either inflation (for example, Becsi and Duca 1994, Duca 1994a, and Koenig 1995b) or nominal GDP (for example, Darin and Hetzel 1994 and Duca 1994b) in the 1990s.[5] Furthermore, these studies find that coefficients on the long-run relationships between an M2-type aggregate and either prices or nominal output change relatively less for the expanded M2 aggregates as samples are extended into the 1990s. From a monetarist perspective, this is an important finding because if velocity is stable in the long run, then monetary aggregates should provide information about medium- to long-run inflationary pressures.[6]

The second type of evidence is that as samples are extended into the 1990s, the impact of asset yields (especially long-term interest rates) varies less in models of money (see Duca 1995 and Koenig 1995a), inflation (see Becsi and Duca 1994, Duca 1994a, and Koenig 1995b), and nominal output (see Duca 1994b) when an expanded M2-type aggregate replaces M2.[7] These findings are consistent with the view that adding bond and stock funds to M2 reduces the omitted variable bias that arises from a lack of data on financial technology and preferences.

The third kind of evidence is cross-section data that confirm a recent shift away from certificates of deposit (CDs) toward bond and equity fund assets (see Kennickell and Starr-McCluer 1994). In particular, during the period 1989–92, when M2 growth was unusually weak, the share of households having nonmoney (mainly bond and equity) mutual fund assets rose from 7.1 to 11.2 percent, whereas the share owning CDs (small and large time deposits) fell from 19.4 to 16.6 percent. Furthermore, over this period, the median value of nonmoney mutual fund assets rose from \$11,200 to \$18,000 among households having such assets, while the median value of CDs rose by a much smaller magnitude—from \$12,600 to \$13,500—among households owning CDs.

While all three types of findings are consistent with the view that models using M2 suffer from omitted variable bias, they do not provide evidence on the actual sources of that bias. The next three sections provide some evidence on these sources.

Technology and shifts toward nonmonetary assets

Since the early 1980s, the attractiveness of nonmonetary assets has likely increased because of two types of technological change: declining costs of transferring funds from nonmonetary assets to transactions deposits and the rising use of financial services from nonasset products.

Lower asset transfer costs. As shown above, a decrease in the cost of shifting funds from savings deposits to transactions accounts (α) and from nonmonetary assets to transactions accounts (β) should reduce the transactions and precautionary demands for money. There are several indications that such costs have fallen. With respect to mutual funds, Orphanides, Reid, and Small (1994) cite evidence that load (commission) fees have fallen sharply over the past two decades. Furthermore, many mutual funds provide customers with a number of free transfers among funds in asset management accounts (see Donoghue Organization 1987) that offer a host of investments, including bonds, equities, and commodities, and allow low- or no-cost shifts among investments within mutual fund families that typically include a checkable money market fund. In addition, many banks now offer mutual funds and have introduced integrated customer management of their mutual fund and bank deposit balances. Additionally, the Federal Reserve has made it easier for households to purchase Treasury securities directly, a change that, coupled with interest rate movements, may have spurred shifts from M2 into Treasury securities, as documented by Feinman and Porter (1992).

More generally, the spread of better information technology is lowering transfer costs, with respect to both domestic and foreign assets (see the box entitled "Globalization"). In particular, the rise of electronic banking (especially using personal computers at home or in the office) poses potentially large reductions in the pecuniary and convenience costs of making such transfers. (For recent evidence, see Holland and Cortese 1995 and Lewis 1995.) Unfortunately, there is no continuous time series of data on asset transfer costs. Nevertheless, the limited evidence is consistent with the fact that most of the unusual weakness in M2 during the 1990s has been concentrated in small time deposits (which compete with stocks and bonds) and money market mutual funds (which experienced outflows when stock and bond yields rose relative to short-term money market rates in the early 1990s).

Financial services from nonassets. Since the 1960s, firms and households have increasingly used new nonasset instruments and cash management techniques to reduce the average level of liquid funds held to meet unexpected cash outflows. In practice, these instruments enable firms and households to better coordinate cash inflow with cash outflow and to reduce check usage by consolidating many purchases into fewer check payments. Within the context of the Milbourne model, these instruments can be interpreted as reducing the volatility of net cash flow (γ) and thereby lowering the demand for money.

In the 1970s and 1980s, technological advances and high interest rates induced firms to seek alternatives to using non-interest-bearing demand deposits to meet their transactions needs.[8] Sophisticated cash management techniques enabled firms to better forecast cash needs and to more readily tap nonmonetary liquid assets to meet unexpected shortfalls in cash flow. (For evidence, see Mahoney 1988 and Porter, Simpson, and Mauskopf 1980, and for more references, Judd and Scadding 1982.) In particular, technological advances spurred many firms to use wire or electronic transfers to minimize transactions balances (see Dotsey 1984 and Flannery and Jaffee 1973).

Although there has been much research on how off-balance-sheet innovations affect the money balances of firms, their effects on household balances have been relatively ignored, even though household transactions balances are larger than those of firms. This lack of research partly reflects that financial innovations spread to households a bit later (in the 1980s and 1990s), after enhancements to computer software whittled down the economies of scale that had made innovations more cost-effective for firms. By providing off-balance-sheet liquidity, the rapid spread of credit cards and credit lines may have enabled households to shift their portfolios away from liquid assets to other assets[9] and may have encouraged households to shift toward risky assets by enabling them to tolerate more price volatility among the assets they hold.

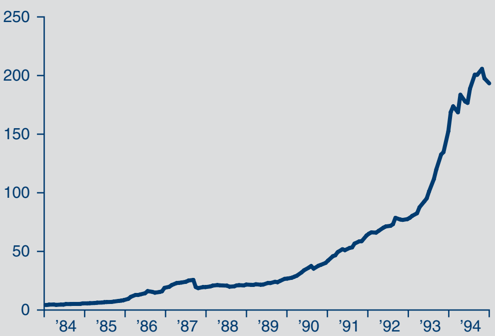
Using cross-section data from 1983, Duca and Whitesell (1995) find that for every 10-percentage-point rise in the probability of owning a credit card, checking accounts are 9 percent smaller, while MMMF plus MMDA balances are 11 percent lower. Although their findings indicate that credit cards significantly affected transactions account levels, they found no statistically significant effect on overall M2 account balances, implying that credit cards primarily affected the composition of M2 in the early 1980s. The impact of credit cards on transactions balances may be even larger today because the share of households owning credit cards is higher, credit cards are more widely accepted, credit card purchases are more quickly processed, and consumers are now offered greater cash rebate/airline mile incentives to use credit cards.[10]

Globalization

Shifts between nonmonetary and monetary assets also involve foreign assets. In terms of the theoretical model, the costs of transferring from foreign to M2 transactions assets (the β for foreign assets) and the risk premium for holding risky foreign assets (*b*) have arguably fallen. Recent studies generally conclude that financial markets across countries have become increasingly integrated (see Obstfeld 1994, 1995 and Feldstein and Bacchetta 1991).[1]

This has manifested itself during the 1990s in at least three ways. First, global bond and equity mutual fund assets have recently grown rapidly (*Figure A*), which may have depressed domestic money holdings and funded overseas activity.[2] Second, financial planners typically recommend that household portfolios contain some foreign assets to improve diversification and to expand the menu of investments having higher expected yields. Third, an enhanced ability to shift funding sources has enabled banks to pull in foreign funds to fund domestic credit growth or use domestic deposits to fund overseas credit growth. In 1993 and 1994, banks pulled in funds from overseas offices—by increasing their net liabilities to foreign offices (*Figure B*)—to fund strong credit growth amid weak M3 growth.[3,4] However, because this funding enables banks to avoid raising deposit rates, it likely restrained M2 and M3 growth in ways captured by the opportunity cost terms in money models.

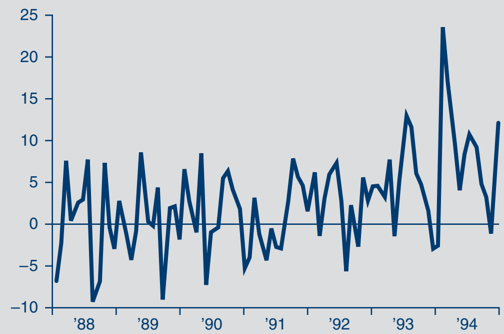
Figure A
International Bond and Equity Mutual Fund Assets*



* Household and institutional assets in mutual funds classified by the Investment Companies Institute as having investment objectives falling under global bond fund, international, or global equity.

SOURCE: Investment Companies Institute.

Figure B
Change in U.S. Banks' Net Liability Position with Related Foreign Offices



1 Obstfeld (1994) finds that marginal rates of substitution in consumption are converging across countries. Feldstein and Bacchetta (1991) find a decline over time in the positive correlation of domestic savings and investment. Both studies find evidence consistent with the view that capital is flowing across borders to areas characterized by relatively higher credit demand.

2 While much, but not necessarily all, of this rise reflects substitution between domestic bonds and equity and foreign bonds and equity, some likely reflects shifts between domestic transactions deposits and foreign securities, consistent with a decline in transfer costs and an apparent decline in the risk premium demanded by U.S. residents to hold these foreign assets.

3 This is measured by the net extent to which commercial banks in the United States are borrowing funds from related foreign offices. The data plotted are from the liability category, "net due to related foreign offices," in the Federal Reserve's H.8 data release.

4 A related, earlier phenomenon was the rise of Eurodollars (offshore dollar-denominated bank deposits) in the late 1970s and early 1980s (Tinsley, Garrett, and Friar 1981), which prompted the inclusion of overnight Eurodollars in M2 and of term Eurodollars in M3.

Another important innovation is the spread of automatic teller machines (ATMs), which reduce the need to carry precautionary currency balances by enabling households to shift nontransactions M2 deposits into cash or transactions accounts. In terms of the theoretical model, ATMs plausibly lower α, the cost of transferring assets within M2, and should thereby lower holdings of transactions deposits and total M2 deposits, with a larger effect on transactions deposits in percentage terms.[11] Using cross-section data from the 1984 and 1986 Surveys of Currency and Transaction Account Usage, Daniels and Murphy (1994a) find that a 100-percentage-point rise in the probability of ATM use increased the velocity of currency (the dollar volume of transactions divided by currency) by 40–45 percent for transactions account holders, while Daniels and Murphy (1994b) estimate that a 5-percent rise in the proportion of ATM users (from 41.7 to 43.8 percent) would boost average transactions account balances by 4.5 percent. Together, these studies imply that ATMs induced households to shift from holding cash to holding transactions account balances in the mid-1980s. Unfortunately, Daniels and Murphy (1994a, 1994b) do not estimate the effect of ATMs on currency plus transactions balances, which corresponds to transactable funds (*D*) in the Milbourne model.

Evidence shows that household payments innovations affected the composition of M2 in the early to mid-1980s. However, the costs of shifting from nonmonetary to transactions assets has fallen since then. Together, lower transfer costs and greater use of nonmoney payments media could now be lowering M2, in addition to altering its composition.[12] For example, many mutual funds offer credit lines and cards with asset management accounts.

The possible roles of demographics, preferences, and learning

Consumer demand theory implies that changes in attitudes toward risk can affect the asset allocations of households. Some of these changes can arise from shifting demographics and economic factors that lead to increased financial sophistication or greater tolerance for investment risk.

Demographics. According to the life-cycle theory of consumption, households save more in their peak earning years before retirement. This theory implies that as the baby-boom generation reaches middle age, the overall savings rate and the portfolio share of higher earning nontransactions assets should rise.

In terms of the Milbourne model, these effects can be accounted for in two possible ways. First, demographic trends may, by increasing the average need to provide for retirement, plausibly raise the willingness of households to invest in risky assets with higher expected long-term yields. In terms of the theoretical framework presented earlier in this article, a lower average degree of risk aversion is reflected in a smaller value of the parameter *b*. This, in turn, raises the risk-adjusted opportunity cost of money for a given spread between the return on nonmonetary assets and money[13] and thereby reduces the demand for money. Alternatively, as people reach their peak earning years, their ratio of income to transactions falls. In terms of the Milbourne model, M2 holdings decline because of a permanent negative shock to ϵ that reduces the demand for money at each combination of income, asset transfer costs, net cash flow volatility, and opportunity cost spreads (see equation 8).

While the post-1980 decline in the U.S. savings rate may contradict the life-cycle theory,[14] recent evidence supports its implications for asset portfolios. With respect to M2, Duca and Whitesell (1995) find, using cross-section data from 1983, that M2 holdings—and in particular, small time deposit and savings balances—are higher for older age brackets after controlling for other variables (for example, income and wealth). Finally, Morgan (1994) finds that the share of household assets held in the form of stocks and bonds is positively correlated with the population share of 35- to 54-year-olds.[15]

Changing preferences and financial sophistication. One factor that could be making monetary assets less attractive is households' increased awareness of investment opportunities in nonmonetary assets and an associated rise in their willingness to tolerate risk in the assets they control (that is, *b* is smaller in the theoretical model presented earlier in this article).[16] Aside from the technological reasons for this trend already mentioned, increased uncertainty in labor markets, changing employment patterns, and the liberalization of IRA/401K accounts have resulted in more households having a hand in managing their retirement assets.[17] This, in turn, has induced households to incur large, predominantly one-time costs to learn more about bond and equity investments for retirement. In addition, because IRA/Keogh balances count toward the minimum balance requirements for opening asset management accounts with many mutual funds, these retirement funds reduce the effective minimum balance requirement on non-IRA/Keogh mutual fund assets. Consistent with this, both IRA/Keogh and non-IRA/Keogh assets with bond and equity mutual funds rose in the mid-1980s, after IRA/Keogh tax laws were eased, and in the early 1990s (see Duca 1995). Additionally, cross-section data indicate a general shift in household portfolios toward bond and equity mutual funds regardless of tax status (see Kennickell and Starr-McCluer 1994).

These factors are consistent with a recent study by Blanchard (1993), who found that the extra return that investors demand from equities over bonds (the "equity premium" of Mehra and Prescott 1985) has been trending downward since the 1940s and abruptly fell in the early 1980s. Five factors likely contributed to the decline in the equity premium: (1) the waning effects of the 1929 stock market crash on risk aversion to stock price movements; (2) investors' realization, following the bond market debacle of the 1970s, that bonds also pose price risk; (3) the rising ownership share of equities held by institutional investors, who are less risk averse and more long-term oriented than households; (4) lower costs for equity diversification, as evidenced by the proliferation of diversified, no-load equity funds; and (5) declining risk aversion among individual investors as they accumulated wealth, gained experience in managing their IRA/Keogh assets, and saw the stock market recover from temporary price corrections in October 1987 and October 1989. As a result of a possible decline in risk aversion, investors may have shifted away from low-risk money assets toward nonmonetary assets that pose higher risk. Nevertheless, because the econometric money results presented earlier could arise for other reasons (for example, technological advances), it is difficult to verify whether and to what extent a systematic shift in risk preferences has noticeably affected money holdings.

Volatility in financial transactions

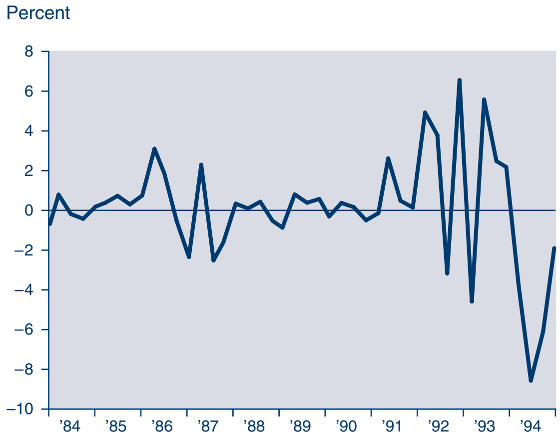
The impact of financial churning on money holdings is more transparent when one recalls that the quantity theory of money implies a relationship between money and transactions, rather than between money and output:

(10) $V = (P \times T)/M = [P \times Y \times (1 + \epsilon)]/M.$

In practice, many analysts implicitly or explicitly replace T with the level of production or consumption of goods and services and redefine velocity accordingly. If the ratio of output to total transactions is stable or predictable,[18] then this substitution does not result in any significant errors in predicting near-term nominal GDP. However, if a monetary aggregate is unusually affected by volatility in non-output transactions, then that aggregate may give a false signal about nominal output. Two recent sources of such volatility have been mortgage refinancings and overseas use of currency.

Mortgage refinancings. In practice, volatility in commercial financial transactions has not affected the average monthly levels of monetary aggregates much in the recent past. One major reason is that economies of scale allow many firms to use wire or electronic transfers to shift funds from nonmonetary assets to settlement funds without having to hold large money balances for a noticeable period of time. While many well-off households similarly manage their mutual fund balances, the way funds are transferred when households prepay mortgages underlying mortgage-backed securities (MBSs) has had large effects on demand deposits, which constitute a large share of M1 and a smaller share of M2.

Figure 1
Estimated Effect of MBS Refinancings On Demand Deposit Growth (Seasonally Adjusted Annual Rate)



SOURCES: Federal Reserve Board; Federal National Mortgage Association; Government National Mortgage Association.

These prepayment effects arise because the Government National Mortgage Association effectively requires MBS servicers to place funds from unscheduled repayments into demand deposit accounts until the fifteenth of the following month before they make principal payments to MBS holders. The Federal National Mortgage Association (FNMA) requires that prepayments be put into custodial accounts until the nineteenth of the following month. While FNMA servicers are not required to put such funds in demand deposit accounts, many do. Because the MBS market was relatively undeveloped until the early to mid-1980s, these effects have only occurred during the mortgage refinancing booms of the mid-1980s and early 1990s. Duca (1990) estimates that swings in MBS prepayments coupled with other, less important effects accounted for one-third of the demand deposit errors from a Federal Reserve econometric model over 1986:1–88:2. Using his methodology, these effects on demand deposits were larger in the early 1990s (*Figure 1*), with estimated effects ranging from adding 6 percentage points to the annualized growth rate in fourth-quarter 1992 to subtracting nearly 9 percentage points in second-quarter 1994. Unless practices change, waves of refinancing activity will likely distort monthly growth patterns of demand deposits and other transactions deposits (see Anderson 1993).

Foreign use of U.S. currency. Fluctuations in the share of currency that is held abroad also distorts growth in narrow measures of money, such as M1 and the monetary base (currency plus reserves). According to reports, use of the dollar has surged in countries suffering from high inflation and political uncertainty. If true, then much of the recent movement in the currency component of U.S. money measures may reflect foreign, rather than domestic, nominal economic activity.

Conclusion

Recent studies generally conclude that the link between nominal output, interest rates, and conventional definitions of broad money has weakened or shifted. By reviewing the recent literature in the context of a microtheoretic model of money, this article attempts to shed light on why these relationships have changed. Three basic factors that may have caused this instability are identified: volatility in financial transactions, technological changes affecting expected transfer costs, and shifts in preferences or demographics that have altered household risk tolerance. In general, while volatility in financial transactions has had substantial effects on narrow monetary aggregates (M1 or the monetary base), it has not been a major source of instability for the broader aggregates. Most of the recent instability in M2's link to nominal GDP does not stem from temporary financial churning or excessive short-term volatility but, rather, reflects an underlying shift in longer term relationships.

By contrast, there is increasing evidence that technological innovations have allowed households to shift away from narrow money or M2 assets toward other financial assets either by reducing asset transfer costs or by allowing households to obtain liquidity via credit lines or electronic transfers. Changing preferences and demographic factors may also be heightening the extent to which other financial assets substitute for money, as manifested by an apparently greater tolerance for risk-taking and a growing share of households that invest their retirement assets.

Changes in technology, and possibly preferences, may continue to alter the relationships between monetary aggregates and nominal variables in coming years. The information revolution will likely foster the spread of electronic financial management, which will further lower asset transfer costs, reduce the need to hold transactable assets in order to obtain liquidity, and lower information barriers that discourage the holding of non-M2 financial assets. In tandem with information advances, greater job mobility, changing employment patterns, and tax incentives are likely to continue bolstering households' role in managing their retirement assets. This greater investment role may, in turn, continue to make households more willing to consider investment alternatives to conventionally defined "money."

Notes

I thank Michelle Thomas for research assistance; Ken Emery, Joe Haslag, and Evan Koenig for comments and suggestions; and the late Stephen Goldfeld and my many colleagues throughout the Federal Reserve System for sharing their insights on money with me over the years.

1 Waud's (1975) model synthesizes Tobin's (1958) portfolio approach with the cash management insights of Miller and Orr (1966). Milbourne's (1986) model is used in this article because it is relatively more transparent. Milbourne's model is used rather than that of Baumol (1952) and Tobin (1956) for two reasons. First, the Milbourne framework can be used to analyze shifts between nontransactions M2 deposits and non-M2 assets, whereas the Baumol–Tobin framework is a model of transactions balances. Second, unlike the Baumol–Tobin model, the Milbourne model allows for uncertainty in cash flow that plausibly affects households' precautionary demand for money.

2 This follows from the fact that $\partial M2/\partial\beta$ and $\partial M2/\partial\alpha > 0$.

3 This follows from the fact that $\partial M2/\partial b > 0$.

4 If innovations primarily lower β and thereby induce shifts between savings deposits and non-M2 assets, then Milbourne's model implies that one should put more emphasis on more narrowly defined money measures that are not affected by such shifts. Nevertheless, even narrow money measures remain vulnerable to innovations, especially given demand shifts that occurred in the early 1980s.

5 Orphanides, Reid, and Small (1994) come to a different conclusion, but their econometric models omit information from the long-run relationship (cointegrating vector) between money and output, in contrast to Duca's model (1994b).

6 This was one of the main motivations for the development of the P* model of Hallman, Porter, and Small (1991).

7 Feinman and Porter (1992) also find evidence that M2's sensitivity to long-term interest rates has risen since the late 1980s.

8 The innovations induced by high interest rates are an example of Lucas's (1976) argument that behavior is not invariant to policy (the Lucas Critique). For a theoretical model of endogenous monetary innovation, see Ireland (1995).

9 Many credit cards enable a household to consolidate

the settlement of many transactions into one monthly payment that has an interest-free grace period. A household can thus lower its *average* liquid deposit balance by making one monthly transfer or by depositing a paycheck before a credit card bill is due.

[10] See Whitesell (1992) for an analysis of how relative costs of using cash, checks, and credit cards affect the use of different payment media.

[11] A decline in α reduces a household's need to hold transactions deposits. Since non-M2 assets have higher pecuniary yields than M2 savings deposits, a decline in α does not induce a rise in savings deposits that offsets the decline in transactions balances.

[12] Research on this issue is currently under way.

[13] Recall that the opportunity cost of money is $[E(r_b) - bvar_{rb} - r_m]$ for transactions accounts in M2 and $[E(r_b) - bvar_{rb} - r_s]$ for nontransactions accounts.

[14] There is much controversy over whether savings behavior supports the life-cycle and permanent income hypotheses. Some, such as Carroll (1992) and Carroll and Kimball (1995), argue that labor income uncertainty limits how far ahead households plan, implying that saving for retirement is much lower than the certainty versions of these theories imply. Others, such as Feldstein (1995a, 1995b, 1974), argue that private pensions, Social Security, and college financial aid programs discourage saving; by implication, the depressing impact of social insurance programs on savings may offset any boost from demographic effects.

[15] Other evidence contradicts Morgan's hypothesis that the aging of the baby boomers accounts for the missing money of the early 1990s. First, the decline in the population share of 35- to 54-year-olds during the early 1970s was not accompanied by unusually strong money growth but, rather, by the first case of the missing money. In addition, aging effects alone cannot account for why money models typically find that M2's sensitivity to long-term interest rates has risen since the 1980s. Finally, the stock and bond market busts of the 1970s may account for the low portfolio share of these securities in that decade, while the higher portfolio shares seen since the mid-1980s may reflect other factors, such as the mid-1980s liberalization of IRAs and Keoghs (see Duca 1995), stronger bond and equity markets since the early 1980s, and a fall in the risk premium on equities (see Blanchard 1993).

[16] While Friedman (1995) points out that households are typically more risk averse than traditional pension fund managers in investing retirement assets, the experience of directing the investment of retirement assets has likely made many people more tolerant of risk for the investments they control.

[17] Since the 1970s, there has been a shift away from defined benefit pension plans toward defined contribution pension plans. One advantage of defined contribution plans is that a greater share of the expected benefits is portable if employment at a particular firm ends.

Gustman and Steinmeier (1992) and Ippolito (1995) estimate that half of the rise in the share of defined contribution plans (401K and traditional defined contribution plans as a share of primary pension plans) owes to employment shifts away from firms that historically have favored defined benefit plans—particularly unionized and larger firms. Ippolito (1995) concludes that the other half of this rise stems from tax law changes that made 401K plans more attractive than pre-1980 defined contribution plans.

[18] For example, if the ratio predictably declines with time, then one can include time trends in predicting velocity (V) and then back out a forecast of nominal output.

References

Anderson, Richard G. (1993), "The Impact of Mortgage Refinancing on Money Demand and the Monetary Aggregates," Federal Reserve Bank of St. Louis *Review,* July/August, 49–63.

Baumol, William (1952), "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quarterly Journal of Economics* 66 (November): 545–56.

Becsi, Zsolt, and John V. Duca (1994), "Adding Bond Funds to M2 in the P-Star Model of Inflation," *Economics Letters* 46 (October): 143–47.

Blanchard, Olivier J. (1993), "Movements in the Equity Premium," *Brookings Papers on Economic Activity*, no. 2: 75–138.

Carroll, Christopher (1992), "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, no. 2: 61–156.

———, and Miles S. Kimball (1995), "On the Concavity of the Consumption Function," manuscript, Board of Governors of the Federal Reserve System (Washington, D.C., February).

Collins, Sean, and Cheryl L. Edwards (1994), "Redefining M2 to Include Bond and Equity Mutual Funds," Federal Reserve Bank of St. Louis *Review,* November/December, 7–30.

Daniels, Kenneth N., and Neil B. Murphy (1994a), "The Impact of Technological Change on the Currency Behavior of Households: An Empirical Cross-Section Study," *Journal of Money, Credit, and Banking* 26 (November): 867–74.

———, and ——— (1994b), "The Impact of Technological Change on Transactions Account Balances: An Empirical Cross-Section Study," *Journal of Financial Services Research* 17 (January): 113–19.

Darin, Robert, and Robert L. Hetzel (1994), "A Shift-Adjusted M2 Indicator for Monetary Policy," Federal Reserve Bank of Richmond *Economic Quarterly,* Summer, 25–48.

The Donoghue Organization (1987), *Donoghue's Mutual Funds Almanac, 1987–1988* (Holliston, Mass.: Donoghues).

Dotsey, Michael (1984), "An Investigation of Cash Management Practices and Their Effects on the Demand for Money," Federal Reserve Bank of Richmond *Economic Review*, September/October, 3–12.

Duca, John V. (1995), "Should Bond Funds Be Included in M2?" *Journal of Banking and Finance* 19 (April): 131–52.

——— (1994a), "Commentary on Mutual Funds and Monetary Policy," Federal Reserve Bank of St. Louis *Review*, November/December, 67–70.

——— (1994b), "Would the Addition of Bond or Equity Funds Make M2 a Better Indicator of Nominal GDP?" Federal Reserve Bank of Dallas *Economic Review*, Fourth Quarter, 1–14.

——— (1990), "The Impact of Mortgage Activity on Recent Demand Deposit Growth," *Economics Letters* 32 (February): 157–61.

———, and William C. Whitesell (1995), "Credit Cards and Money Demand: A Cross-Sectional Study," *Journal of Money, Credit, and Banking* 27 (May): 604–23.

Feinman, Joshua, and Richard D. Porter (1992), "The Continued Weakness in M2," FEDS Working Paper no. 209, Board of Governors of the Federal Reserve System (Washington, D.C., September).

Feldstein, Martin (1995a), "College Scholarship Rules and Private Saving," *American Economic Review* 85 (June): 552–66.

——— (1995b), "Social Security and Saving: New Time Series Evidence," NBER Working Paper Series, no. 5054 (Cambridge, Mass.: National Bureau of Economic Research, March).

——— (1974), "Social Security, Induced Retirement and Aggregate Capital Accumulation," *Journal of Political Economy* 82 (September/October): 905–26.

———, and Phillipe Bacchetta (1991), "National Saving and International Investment," in *National Saving and Economic Performance*, ed. B. Douglas Bernheim and John B. Shoven (Chicago: University of Chicago Press).

Flannery, Mark J., and Dwight M. Jaffee (1973), *The Economic Implications of an Electronic Monetary Transfer System* (Lexington, Mass.: Lexington Books).

Friedman, Benjamin M. (1995), "Economic Implications of Changing Share Ownership," NBER Working Paper Series, no. 5141 (Cambridge, Mass.: National Bureau of Economic Research, June).

Friend, Irwin, and Marshall E. Blume (1975), "The Demand for Risky Assets," *American Economic Review* 65 (December): 900–22.

Goldfeld, Stephen M. (1976), "The Case of the Missing Money," *Brookings Papers on Economic Activity*, no. 3: 683–730.

——— (1973), "The Demand for Money Revisited," *Brookings Papers on Economic Activity*, no. 3: 577–638.

Gustman, Alan L., and Thomas L. Steinmeier (1992), "The Stampede Toward Defined Contribution Pension Plans: Fact or Fiction?" *Industrial Relations* 31 (Spring): 361–69.

Hallman, Jeffrey J., Richard D. Porter, and David H. Small (1991), "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?" *American Economic Review* 81 (September): 841–58.

Hess, Gregory D., and Charles Morris (1995), "Money Is What Money Predicts: The M* Model of the Price Level," Federal Reserve Bank of Kansas City Working Paper no. 9505 (Kansas City, June).

Hetzel, Robert L., and Yash P. Mehra (1989), "The Behavior of Money Demand in the 1980s," *Journal of Money, Credit, and Banking* 21 (November): 455–63.

Holland, Kelley, and Amy Cortese (1995), "The Future of Money," *Business Week,* June 12, 66–78.

Ippolito, Richard A. (1995), "Toward Explaining the Growth of Defined Contribution Plans," *Industrial Relations* 34 (January): 1–20.

Ireland, Peter (1995), "Endogenous Financial Innovation and the Demand for Money," *Journal of Money, Credit, and Banking* (February): 107–23.

Judd, John P., and John L. Scadding (1982), "The Search for a Stable Money Demand Function: A Survey of the Post-1973 Literature," *Journal of Economic Literature* 20 (September): 993–1023.

Kennickell, Arthur B., and Martha Starr-McCluer (1994), "Changes in Family Finances from 1989 to 1992: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, October, 861–82.

Koenig, Evan F. (1995a), "Long-Term Interest Rates and the Recent Weakness in M2," manuscript, Federal Reserve Bank of Dallas (Dallas, June).

——— (1995b), "The P* Model of Inflation Revisited," manuscript, Federal Reserve Bank of Dallas (Dallas, February).

Lewis, Peter H. (1995), "Chemical Aims to Expand Electronic Banking," *New York Times,* July 7, D5.

Lucas, Robert E., Jr. (1976), "Econometric Policy Evaluation: A Critique," *Carnegie–Rochester Conference Series on Public Policy* 1: 19–46.

Mahoney, Patrick I. (1988), "The Recent Behavior of Demand Deposits," *Federal Reserve Bulletin*, April, 195–208.

Mehra, Rajnish, and Edward C. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15 (March): 145–61.

Milbourne, Ross (1986), "Financial Innovation and the Demand for Liquid Assets," *Journal of Money, Credit, and Banking* 18 (November): 506–11.

Miller, Merton H., and Daniel Orr (1966), "A Model of the Demand for Money by Firms," *Quarterly Journal of Economics* 80 (August): 413–35.

Moore, George R., Richard D. Porter, and David H. Small (1990), "Modeling the Disaggregated Demands for M1 and M2 in the 1980's: The U.S. Experience," in *Financial Sectors in Open Economies: Empirical Analysis and Policy Issues*, ed. P. Hooper, K. H. Johnson, D. L. Kohn, D. E. Lindsey, R. D. Porter, and R. Tryon (Washington, D.C.: Board of Governors of the Federal Reserve System), 21–105.

Morgan, Donald P. (1994), "Will the Shift to Stocks and Bonds by Households Be Destabilizing?" Federal Reserve Bank of Kansas City *Economic Review*, Second Quarter, 31–44.

Obstfeld, Maurice (1995), "Capital Mobility in the 1990s," International Finance Discussion Paper no. 472 (Washington, D.C.: Board of Governors of the Federal Reserve System, June).

——— (1994), "Are Industrial-Country Consumption Risks Globally Diversified?" in *Capital Mobility: The Impact on Consumption, Investment, and Growth*, ed. Leonardo Liederman and Assaf Razin (Cambridge: Cambridge University Press).

Orphanides, Athanasios, Brian Reid, and David H. Small (1994), "Empirical Properties of a Monetary Aggregate That Adds Bond and Stock Funds to M2," Federal Reserve Bank of St. Louis *Review*, November/December, 31–52.

Porter, Richard D., Thomas D. Simpson, and Eileen Mauskopf (1980), "Financial Innovation and the Monetary Aggregates," *Brookings Papers on Economic Activity*, no. 1: 213–29.

Simpson, Thomas D. (1980), "The Redefined Monetary Aggregates," *Federal Reserve Bulletin*, February, 97–114.

Small, David H., and Richard D. Porter (1989), "Understanding the Behavior of M2 and V2," *Federal Reserve Bulletin*, April, 244–54.

Tinsley, P. A., B. Garrett, and M. E. Friar (1981), "An Exposé of Disguised Deposits," *Journal of Econometrics* 15 (January): 117–37.

Tobin, James (1958), "Liquidity Preference as Behavior Toward Risk," *Review of Economic Studies* 25 (February): 65–86.

——— (1956), "The Interest Elasticity of Transactions Demand for Cash," *Review of Economics and Statistics* 38 (August): 241–47.

Waud, Roger N. (1975), "Net Outlay Uncertainty and Liquidity Preference as Behavior Toward Risk," *Journal of Money, Credit, and Banking* 7 (November): 499–506.

Whitesell, William C. (1992), "Deposit Banks and the Market for Payment Media," *Journal of Money, Credit, and Banking* 24 (November): 483–98.

Argentina, Mexico, And Currency Boards: Another Case of Rules Versus Discretion

Carlos E. Zarazaga
Senior Economist and Executive Director
of the Center for Latin American Economics
Federal Reserve Bank of Dallas

Far more important than what governments say—or even enact into law—seems to be what governments do: actions speak louder than words or laws. A country's care for its reputation plays a far more important role than formal institutions in solving the time inconsistency problem and in providing governments with the incentives to adhere to policy rules despite the short-term temptation to do otherwise.

In years to come, world financial markets will recall December 20, 1994, the day Mexico devalued its currency, as a landmark date in financial history. The devaluation inadvertently initiated what Michel Camdessus, managing director of the International Monetary Fund, aptly dubbed "the first financial crisis of the twenty-first century." Most analysts and economic advisors were surprised by not only the devaluation, but also the speed with which its effects spilled into other emerging economies. These effects took the form of swift and massive capital outflows, as investors withdrew savings from those countries in fear that they would devalue their currencies as well.

The *tequila effect*, as the Mexican crisis has come to be known in Latin America, has eroded the living standards of millions of people throughout the region.[1] Also affected, although to a lesser extent, are the countries and international organizations that quickly assisted Mexico with an unusual financial package. The damaging effects, actual or potential, of the Mexican crisis on so many people's welfare have caused the public, investors, and the press to question how Mexico's crisis happened, how it influenced other economies, and how to prevent a similar crisis in the future. The response of some analysts has been that the Mexican crisis and its daunting spillover effects would have been avoided had Mexico had a currency board-like system similar to the one Argentina adopted in 1991.

My goal in this article is to examine the currency board proposition in light of current economic theory and the experiences of Argentina and Mexico. In the first part of the article, I describe the monetary policies of those two countries and argue that Mexico was forced to devalue its currency while Argentina was not because Mexico managed its monetary policy with much more discretion than did Argentina, which managed monetary policy according to strict rules.

The seemingly obvious conclusion of the first part of the article is that all it takes to prevent exchange rate crises such as Mexico's is to guarantee that rules will take precedence over discretion. Currency boards, their advocates maintain, provide governments with the adequate "technology" with which to handle such a simple job.

In the second part of this article, however, I argue that this optimistic view is too naive because it overlooks the problem of time inconsistency,[2] a bit of economics jargon for policymakers' tendency to find good reasons to repudiate plans they had promised not to abandon and policy rules they had vowed not to break. Governments always justify those inconsistencies with the same basic excuse: the abandoned policy rule was the best course in the conditions prevailing in the past but not for present circumstances.[3]

Currency boards are a monetary policy rule. As such they fail to resolve the time inconsistency problem because, despite claims to the contrary, currency boards cannot provide a quick and painless fix to the economic woes of countries that, like Mexico and Argentina, have long histories inconsistent with low-inflation targets. Quite to the contrary, implementation of rules in such countries is bound to be costly because the credibility of each country's economic policies depends more on the country's track record in honoring past commitments than on present institutional arrangements.

In fact, as I argue in the third part of this article, reputation is an important determinant of which rules are best for a country. In general, contingent policy rules or rules with (implicit or explicit) escape clauses are superior to noncontingent rules such as currency boards. But the recent experiences of Argentina and Mexico may suggest that implementation of the more flexible contingent rules is particularly difficult in countries that have inappropriately used in the past built-in escape clauses. By virtue of their poor track records, such countries may be limited to the use of noncontingent rules. Currency boards are one such rule, but certainly not the only one, and policymakers should carefully evaluate the merits and shortcomings of currency boards relative to other types of ironclad rules before recommending currency boards as the best rule for a country.

Whatever rule is chosen, countries that have lacked monetary discipline in the past and attempt to implement strict monetary policies eventually may suffer severe economic hardships. When problems arise, ironclad rules such as currency boards will be particularly susceptible to the time inconsistency problem. Countries will be able to overcome such problems only if their people are convinced that the concrete costs of sticking to the policy rule today will be outweighed by the potential gains that will accrue when investors' confidence is eventually regained. Unfortunately, this cost–benefit analysis is subject to considerably more dispute than currency board advocates sometimes recognize. Nonetheless, this article concludes on the optimistic note that Argentina's and Mexico's recent experiences may provide useful empirical evidence to validate or refute claims about currency boards, principles of time inconsistency literature, and theories about the superiority of rules over discretion.

The monetary policies of Argentina and Mexico

Currency boards: A devaluation-proof rule for money base creation. A currency board is a policy rule for monetary base creation that guarantees that a country will not devalue under any circumstance while following that rule.[4] Under a currency board, monetary policy is run according to a very simple rule: the monetary authority issues money only against a designated reserve currency, such as the U.S. dollar or German mark, at a fixed exchange rate. This rule is formalized in the following equation:

$$(1) \qquad \frac{x}{\textit{Promised Exchange Rate}} = \textit{Stock of Reserve Currency},$$

where x is the level of monetary base that satisfies the equality. In a country that runs its monetary policy according to a currency board rule, all policymakers need to do is print the amount of money that satisfies x in equation 1. This rule implies that if the stock of reserve currency expands by 10 percent (say, due to a capital inflow), then the monetary authority must expand the monetary base by 10 percent. If, in contrast, the stock of reserve currency shrinks by 10 percent (say, due to a capital outflow), then the monetary authority must contract the monetary base by 10 percent. In other words, a currency board mechanism for expanding and contracting the monetary base ensures that the proportion of monetary base to reserves remains constant at the fixed exchange rate. To see this more formally, define

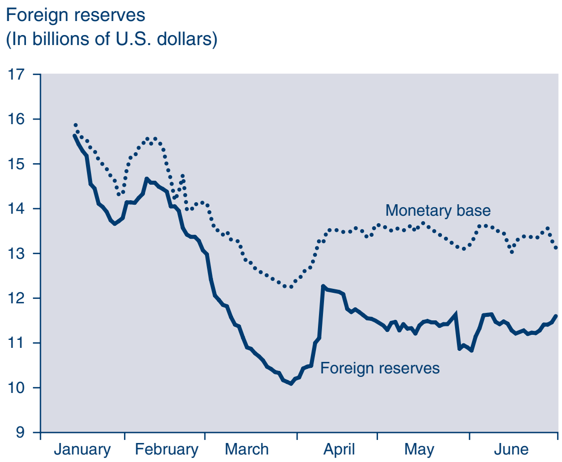
$$(2) \qquad MB\$FR = \frac{\dfrac{\textit{Monetary Base}}{\textit{Promised Exchange Rate}}}{\textit{Stock of Reserve Currency.}}$$

The left-hand term in this equation is the *MB$FR ratio*. A currency board simply instructs the monetary authorities to set that ratio equal to 1, so that

$$(3) \qquad MB\$FR = 1$$

becomes the currency board rule. The economic interpretation of this rule is that the monetary

Figure 1
Argentina: Monetary Base And Foreign Reserves, 1995



SOURCE: Central bank of Argentina.

base is fully backed by the designated foreign reserve currency.

To understand how a currency board works, suppose that for some reason all the households in a country suddenly decide to exchange all the money they have in their country's currency for dollars. Under a currency board system, this massive speculative attack against the local currency will not trigger a devaluation, as it did in Mexico, because a monetary authority adhering to a currency board rule never runs out of the reserve currency and can eventually buy back *all* the monetary base (that is, exchange it for foreign currency) with its reserves at the promised exchange rate.

Unfortunately, currency board advocates often fail to emphasize that the cost of successfully defending the parity between the reserve and domestic currencies may be a severe financial crisis. Lessons about the virtues and shortcomings of a currency board, as well as the events that led to Mexico's peso devaluation, can be drawn from a review of the recent economic experience of Argentina, a country that has been following a quasi-currency board rule very closely since 1991.

Argentina's monetary policy and currency boards. On April 1, 1991, Argentina's congress approved a convertibility law.[5] This law obligates the central bank to issue domestic currency (the peso) almost exclusively against the dollar value of foreign reserves at the fixed exchange rate of 1:1—in other words, at the rate of 1 peso for every $1 received by the central bank. This standard is the basic rule for money creation described in the previous section. Although many policy analysts refer to Argentina's current monetary regime as a currency board, the policy has not been run as an orthodox currency board rule. Even so, the policy so closely resembles a pure currency board regime that it serves as a useful example.

Figure 1 shows the evolution of the monetary base and foreign reserves in Argentina during 1995. On January 1, 1995, the foreign reserves were $15.7 billion, backing a monetary base of 16.3 billion pesos. The *MB$FR* ratio was very close to 1, the ratio stipulated by the currency board rule, so there was practically no difference between a currency board and Argentina's monetary regime on January 1.

If Argentina's policy were a textbook currency board, the two lines in Figure 1 would overlap throughout the figure. The lines do not overlap because, unlike an orthodox currency board, Argentina's convertibility law gives the central bank some flexibility to act as lender of last resort (Zarazaga 1995b). Argentina's central bank can issue money for that purpose up to the level that would push the *MB$FR* ratio above 1.25. Stated differently, the convertibility law does not require 100-percent backing of the monetary base: only 80 percent of it must be backed by foreign reserves (at the committed 1:1 exchange rate).

Had Argentina's policy been a pure currency board, when the country's foreign reserves shrank to about $10 billion in late March 1995, the monetary base would have shrunk by 5.7 billion to 10.6 billion pesos. However, Argentina's monetary base declined only to about 12.3 billion pesos. The *MB$FR* ratio peaked at 1.23 on March 30, 1995.[6] At that time, Argentina's central bank had $1 for every 1.23 pesos of bills and coins in the public's wallets and banks' vaults (or, equivalently, $0.82 for each peso of monetary base). Had the holders of pesos wished to exchange all their cash—the 12.3 billion pesos—for dollars, Argentina would have been forced to devalue its currency by about 23 percent.

Of course, this scenario overstates the risks of a devaluation in Argentina in March 1995 because it would be rare for all individuals and businesses simultaneously to want to rid themselves of the local currency. Some amount, even if modest, of bills and coins will always be needed to carry out transactions such as paying taxes or buying a soda in vending machines. Because some local currency will never be presented in exchange for dollars, the monetary base can grow slightly beyond the currency board limit.

The monetary authority can exploit this fact to manage the monetary base and expand it in moderate amounts, as Argentina's monetary authorities did, to act as lender of last resort. Although such a moderate expansion to help the financial system will not be backed by foreign reserves, the risk of a devaluation will be reasonable if policymakers do not abuse their leeway. Argentina's 80-percent coverage of the monetary base with foreign reserves, for instance, seems prudent.[7]

To summarize, Argentina's quasi-currency board rule has allowed its monetary authorities a little more flexibility in conducting monetary policy than an orthodox currency board would have. Still, Argentina's system imposes very clear limits on discretionary expansions of the monetary base. Monetary authorities respecting similar limits to their discretion within a fixed exchange rate regime will not be able to isolate changes in foreign reserves from changes in the monetary base for too long. Sooner or later, sustained declines in foreign reserves will be reflected in corresponding declines in the monetary base. This is why in Figure 1 the monetary base and foreign reserves in Argentina move in tandem, despite the flexibility built into the country's quasi-currency board regime.

Argentina's quasi-currency board rule under attack. When Argentina's peso came under speculative attack in first-quarter 1995, policymakers could defend the currency because they stubbornly adhered to a policy rule that guaranteed that at least 80 percent of the monetary base would always be covered by foreign reserves. But the price of this success was one of the most severe banking panics in modern Argentine history.

The performance of Argentina's quasi-currency board during a financial crisis illustrates that currency boards can avert devaluations. But because of their very limited ability to act as a lender of last resort, they introduce the risk that a minor, Orange County-type liquidity crisis[8] will become a devastating national financial panic almost overnight.

Argentina's case study demonstrates that currency boards have very little power to control financial crises when they occur in a modern, independent country, rather than in the colonies frequently cited as success stories in the literature of currency board advocates.[9]

Argentina's financial panic started with a liquidity squeeze in Bank Extrader, a small bank that held barely 0.2 percent of the total deposits in Argentina's financial system. Extrader was heavily exposed in Mexican bonds and securities. When the value of those assets fell dramatically in the aftermath of the devaluation of the Mexican peso on December 20, 1994, the bank could no longer cover its short-term liabilities, particularly some time deposits that came due. This shortage triggered a run against the bank. Extrader, unable to honor its deposits, was foreclosed on by the central bank on January 18, 1995.

The fear that other banks were similarly exposed translated into a generalized banking panic. Suddenly, Argentina's financial system was awash in the same indiscriminate chain reaction that had transmitted the tequila effect throughout Latin American capital markets. Almost immediately, the run against the banks became a run against the domestic currency. People feared that Argentina would devalue as Mexico had done shortly before. As depicted by the decline in foreign reserves in Figure 1, much of the cash withdrawn from Argentina's financial system went to purchase dollars that were sent abroad.

By the end of April 1995, Argentina's financial system had lost 18 percent of the deposits it had before the Mexican peso devaluation. As a measure of the severity of this contraction, Argentina experienced in just *three months* the same proportional contraction in deposits as the United States did during the first *two years* of the Great Depression. In the wake of Argentina's financial panic, many banks were forced to suspend the payment of deposits. Many investors—foreign and domestic alike—have yet to recover their savings. Argentina's experience, therefore, should dispel the notion that a currency board would have prevented the financial meltdown Mexico would have suffered without the U.S.–International Monetary Fund aid package.

The complete interruption of the chain of payments and shutdown of credit markets took its toll on Argentina's real economy. Second-quarter gross domestic product (GDP) in 1995 fell by about 5 percent from its second-quarter 1994 level, while the fall in third-quarter 1995 from third-quarter 1994 was 8 percent. These figures have led many private forecasters to conclude that Argentina's 1995 GDP (adjusted for inflation) will be 2.5 percent below that of 1994. Perhaps the most worrisome consequence of the financial crisis was a jump in the country's unemployment rate, from 12.5 percent in October 1994 to an all-time high of 18.6 percent in May 1995.

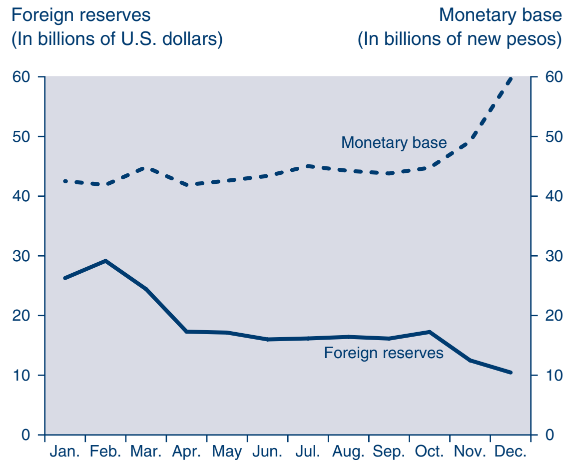
Numbers like Argentina's make it easy to understand why investors may fear countries will abandon currency board-like rules. When countries confront banking crises, such rules provide little more than homeopathic therapy while panics run their natural course.[10] As time inconsistency theory predicts, during times of stress, investors grow skeptical about governments' pledges to honor their commitments to currency board-like rules. Investors conjecture that rising unemployment and eroding political support might force governments to abandon the rule-bound currency board system and replace it with policies prone to devaluation—what the currency board was designed to prevent.[11] That time inconsistency problem is why investors questioned the continuity of Argentina's quasi-currency board rule and why they withdrew their savings from the country. This capital flight, in fact, helped generate the financial crisis that continued the cycle of devaluation fears.

Contrary to the predictions of currency board advocates, the formal legal arrangement of a quasi-currency board did not protect Argentina from a speculative attack against its currency. Argentina's monetary policy, as predicted, prevented a devaluation, but the price was a banking crisis far more severe than currency board advocates had anticipated.

Mexico's discretionary monetary policy. The movement of Argentina's monetary base and foreign reserves displayed in Figure 1 contrasts sharply with that of Mexico's. Figure 2 shows that Mexico's monetary base remained fairly constant and even increased after October 1994, despite a continuous decline in foreign reserves. The difference between the two figures suggests that Argentina was more conservative than Mexico in tolerating deviations from the currency board rule. The *MB$FR* ratio never reached the legal limit of 1.25 in Argentina but was 1.62 in Mexico on December 19, 1994, the day before the devaluation.[12] Undoubtedly, a devaluation is much more likely in a country that backs less than 80 percent of its monetary base with foreign reserves (as Mexico did in late 1994) than in a country that backs 80 percent or more of its monetary base with foreign reserves (as has been the case in Argentina).

Interestingly enough, until October 1994, Mexico had managed its monetary base according to a rule that far exceeded the rigor of the currency board standard. Before fourth-quarter 1994, Mexico's *MB$FR* ratio had been below 1 (that is, Mexico's foreign reserves had exceeded its monetary base). This observation suggests an alternative interpretation of Mexico's monetary policy. Perhaps what differentiated Mexico's experience from Argentina's is not that Argentina passed a law requiring a quasi-currency board

Figure 2
Mexico: Monetary Base And Foreign Reserves, 1994



SOURCE: Banco de México.

rule while Mexico did not but, rather, that sometime after October 1994 Mexico decided to repudiate its policy suddenly and almost without warning.[13]

In any case, Mexico's relatively high fourth-quarter 1994 *MB$FR* ratio implies that Banco de México was no longer in a position to exchange Mexico's entire monetary base for dollars at the promised exchange rate, and that, subject to a speculative attack, Mexico would eventually be forced to devalue its currency by about 60 percent.[14] Unfortunately, the fear of a speculative attack became self-fulfilling and triggered a chain of events that led to the December 20, 1994, devaluation.

But why did the Mexican monetary authorities allow the monetary base to grow without the backing of foreign reserves after October 1994? As Mexican monetary authorities later explained, a continuous drain of foreign reserves had started in February 1994[15] and had exposed the banking sector to the risk of a liquidity crunch. Concerned abut a banking crisis, Mexican monetary authorities tried to preempt a financial panic by acting as a lender of last resort. Discount window loans to allegedly troubled financial institutions expanded the monetary base beyond the level of foreign reserves (Zarazaga 1995b), leaving Mexico vulnerable to speculative attack and devaluation.

On face value, the expansion of Mexico's monetary base through its central bank's discount window—despite declining foreign reserves—may appear inconsistent with the pegged exchange rate regime in place at the time. The action, however, was not necessarily inconsistent, provided the monetary authorities had reasons to believe that the capital outflows and consequent loss of foreign reserves were only temporary and would reverse themselves once the fears of political turmoil subsided,[16] and that the minimum demand for local currency had increased as well.

Banco de México authorities have stated that such reasons did indeed exist,[17] even if now, in hindsight, it may appear that the effects of political uncertainty on Mexico's creditworthiness were underestimated[18] and the increase in demand for Mexican pesos was overestimated.[19]

Undoubtedly, something went wrong. Mexico most likely suffered the same problem that has hit many other countries when their currencies have been devalued after their policymakers miscalculated the leeway they had for expansions of the money base not backed by foreign reserves. In the attempt to fine-tune the economy, even the most skilled policymakers may read the tea leaves incorrectly from time to time. In Mexico's case, the monetary authorities validated the use of the discount window—and, therefore, the increase of the unbacked monetary supply—to a level that, ex post, exceeded what the market was willing to absorb.[20]

If the source of the problem is not necessarily unskilled policymakers but the discretion they enjoy in conducting monetary policy (for example, to preempt bank runs), then the obvious fix would be to take away policymakers' discretion. This is the reasoning behind many enthusiastic recommendations in favor of currency boards and the focus of the next section.

Can institutions eradicate discretionary policies?

Since a currency board is nothing but a rule for money creation, the debate about the advantages, disadvantages, and desirability of currency boards amounts to another rendition of the long-standing rules-versus-discretion debate. Currency board advocates maintain that the Mexican crisis would have been avoided if a currency board like Argentina's had limited the discretion of Mexico's monetary authorities. Although this argument might ring true, it naively attributes to formal rules and institutions more power than they have in committing governments to keep their promises in the face of adverse economic conditions.

The problem is that policy rules, however institutionalized, are inherently time inconsistent—in other words, governments will tend to abandon them. Advocates of currency boards have failed to show how such institutions can overcome this problem. As mentioned earlier, Argentina, despite its quasi-currency board, suffered a speculative attack driven by distrust in the continuity of its monetary policy.

Argentina's example further indicates that legal institutions per se provide very little reassurance about a country's future economic policies. In fact, during Argentina's financial crisis, Art. 17 of that country's central bank charter was modified by presidential decree to give that institution more flexibility in its discount window policies. That charter, approved by law number 24,144 of September 23, 1992, had enacted the central bank independence. But the presidential decree raised and justified the fears that the whole central bank charter and, therefore, central bank independence, would be repudiated. Another indication of how ineffective formal institutions and legal arrangements are in limiting policymakers' discretion comes from German history. The Reichsbank, the central bank of the German Empire, was declared legally independent on May 26, 1922, just three months before the onset of the 1922–23 German hyperinflation.[21]

Besides, neither Germany nor the United States has an explicit or legislated rule for running monetary policy such as Argentina's, yet Germany's or the United States' credibility in keeping inflation low and its currency stable far exceeds Argentina's because Germany and the United States have strong track records.

Far more important than what governments say—or even enact into law—seems to be what governments do: actions speak louder than words or laws. A country's care for its reputation plays a far more important role than its institutions in solving the time inconsistency problem and in providing governments with the incentives to adhere to policy rules despite the short-term temptation to do otherwise. This is the basic insight of Barro and Gordon (1983) and the literature that followed.[22] The credibility of policymakers and economic policies will be much higher in countries with a long tradition of respecting policy rules than in countries with a tradition of repudiating them.[23]

Given the role of reputation, new policy rules will meet considerable skepticism in countries that have failed to demonstrate past discipline. Guided by a country's history of repeatedly broken commitments, economic agents will (justifiably) bet against policy continuity, whether the government promises come in the form of public statements or formal institutions such as currency boards.[24]

Formal institutions or laws cannot remove skepticism about governments' ability to carry out commitments in countries that have repeatedly failed to honor past promises. The adoption of rules in such countries, however implemented, sooner or later is likely to produce severe economic and social hardships while the country persuades investors that it has mended its ways and will no longer abandon its commitments.

Are currency boards the best rule?

The failure to explain how currency boards solve the time inconsistency problem is not the only wrinkle in arguments that portray currency boards as the instant recipe for restored credibility and prosperity. But setting aside the issue of time inconsistency, there is the normative question of which is the best rule. What the literature has established is that optimal rules are superior to discretion.[25] A vast array of plausible policy rules and, in particular, of monetary policy rules is available to policymakers, and economists have yet to reach a consensus that currency board rules are superior to any other feasible rule.

Furthermore, many economists would argue that contingent rules are superior to ironclad ones that are invariant to changing economic contingencies. Several studies, in fact, show that rules with escape clauses are the best course of action.[26]

In this spirit, Bordo and Kydland (1995) argue that, despite appearing to be an ironclad rule, the gold standard in reality had implicit escape clauses. Bordo and Kydland point to periods when England, the country that most consistently adhered to the rule, temporarily suspended convertibility of the pound into gold (at a fixed exchange rate of £3.85 per ounce) during wars and financial crises.[27]

Admittedly, the use of rules with escape clauses opens a Pandora's box because rules with too many contingencies and escape clauses can become indistinguishable from discretion.[28] For example, did Mexico repudiate the fixed exchange rate rule through its extensive lender-of-last-resort activity just before the devaluation? Or, was Mexico simply exercising an escape clause to avert a financial crisis in the face of adverse and unforeseen political shocks, as England did to quench the incipient banking panics of 1847, 1857, and 1866? This will be the subject of considerable debate for many years to come, in part because several empirical and theoretical issues involved remain largely unresolved.[29]

But events in Mexico suggest that financial markets participants did not view the monetary policy actions at the end of 1994 as a temporary and justifiable use of an escape clause. Rather, the markets seem to have confused those policies with superficially similar ones that several years earlier (in 1982 and 1987) had led to devaluations accompanied or immediately followed by violations of elementary free market rules, such as nationalization of banks, confiscation of deposits, open or disguised forms of price and capital controls, and outright default on government debt. As in the tale of the boy who falsely cried wolf too often, Mexican policymakers in 1994 were trapped by the bad reputation of their predecessors.[30]

Perhaps one of the more important lessons of the Mexican crisis of 1994–95 is that the invocation of escape clauses might be unwise in countries that, in the eyes of investors, have abused such outs in the past.[31] For these countries, ironclad rules might well be the only hope to restore investors' confidence and, therefore, future prosperity. But this essentially sound point will be perhaps better served by the recognition that a currency board is just one type of ironclad rule, not necessarily, and certainly not in general, the best one.

Whichever ironclad rule proves best, it is necessary to revisit the issue of how it will overcome the time inconsistency problem. More concretely, will societies accept the immediate costs of implementing a rule, particularly severe in countries with a poor reputation, on the promise of the benefits that will accrue in time?

Minimizing or dismissing the costs of a particular ironclad rule in the zeal of promoting its adoption (as has often been done) could prove self-defeating because a society may too easily become disenchanted and abandon the rule at the first setback, before the rule has had time to take hold and produce the desired results. To the contrary, the cause of rules would be better served if scholars, decisionmakers, and opinionmakers clearly explained to societies the nature of the inevitable economic hardships the rules will entail after years of inconsistent monetary policy.

In this sense, Argentina's decision to respect the quasi-currency board rule despite its serious financial crisis is almost unprecedented. Perhaps Argentina's authorities (and Argentina's people, who reelected the government in the middle of the crisis) were motivated to stick to their guns by a conviction that the alternative, to abandon the currency board, would have been perceived, as in Mexico, not as the use of an escape clause to control a banking panic but as a return to the old ways of running monetary policy. Such past policies were based on almost unbounded discretion and led to decades of impoverishing inflationary stagnation and to a traumatic hyperinflation during 1989–90.[32]

In any case, much of the difficulty policymakers face in choosing among different policy rules arises because the theory of costs and benefits of alternative policies is still well ahead of the empirical evidence available to measure them. For all their catastrophic dimensions, one potentially positive outcome of recent events in Mexico and Argentina might be to help close the theory–evidence gap. After all, those experiences are as close as economists get to controlled experiments needed to measure the costs and benefits of alternative policies: both are Latin American countries with similar characteristics and past histories, and each responded with a different policy to basically the same speculative attack against its currency. Of course, the task of identifying the effects of the different policies followed by those countries so far will not be as easy as the highly stylized, stark identifying assumptions just mentioned might suggest.

There are a number of other important factors that now or in the future could affect the economic outcomes of those two countries. But in economics, as in any other social science, the only feasible experiment is complex, sometimes fuzzy historical evidence, and few economists would argue that we have not learned anything from examining the past. Just the opposite is true, as few economists can resist the temptation of presenting data, which is information from the past, to back up their arguments and theories. It does not seem preposterous, therefore, to think that clever economists will be able to design appropriate quantitative methods to identify and measure cause–effect relationships between the eventually different economic performances and the so far certainly different policy responses of Argentina and Mexico. For that reason, the recent experiences of those two countries are already proving to be a popular and fertile area of research, one that might help assess the wisdom of Argentina's decision to stick to its quasi-currency board arrangement and, in any event, enrich and change the terms of the rules-versus-discretion debate for years to come.

Conclusions

This analysis of the monetary policies of Argentina and Mexico has shown that, unlike Mexico, Argentina prevented a devaluation of its currency by following a quasi-currency board rule. Based on this observation, many have recommended a currency board for Mexico as well. This recommendation, however, is based on the naive belief that the formal institution of a currency board provides a commitment technology that ensures policymakers will conduct monetary policy according to a very well-defined rule.

The truth is that currency boards and similar institutions cannot enforce a government's everlasting commitment to low inflation and pegged or fixed exchange rate policies any more than a wedding ring can ensure a spouse's commitment to an everlasting marriage. This weakness is common to other institutions and written laws as well, and its source is the same: ironclad rules *do not* resolve the basic problem of time inconsistency. This problem lies at the heart of the lack of credibility that haunts policymakers in countries that have frequently broken their commitments in the past. This lack of credibility explains why currency boards are subject to speculative attacks that they can resist without devaluing only at the cost of very severe financial crises.

Therefore, depictions of currency boards—or any other ironclad rule, for that matter—as powerful devices that will magically restore investors' confidence and, therefore, prosperity almost overnight and without pain do not help. On the contrary, this optimistic assessment may have the perverse effect of providing policymakers with the incentive to abandon their commitments on the mistaken impression that later, simply by institutionalizing a rule such as a currency board, they can quickly and painlessly restore lost credibility.

In truth, a government's credibility is like crystal: once broken, it is very difficult and costly to restore. Rules would, perhaps, stand a better chance of overcoming the time inconsistency problem if the governments and societies of countries that abandoned past promises understood the true cost of regaining credibility. The costs of following a sensible monetary rule are the price to pay for the bad reputation that stems from a past of broken trust and for the future economic development that regaining credibility will eventually bring about.

Unfortunately, economic theory has made little progress in predicting when and why countries will finally abandon discretionary policies and switch to rules, or, equivalently, when countries will perceive that future benefits of restored investor confidence outweigh the present economic hardships of rebuilding reputations.

In any case, societies considering commitment to a rule should consider that noncontingent policy rules such as currency boards are, in general, inferior to contingent rules. But because the distinction between pure discretion and contingent rules may become blurred in countries that have abused the flexibility provided by rules with escape clauses, such countries may have pushed themselves into an all or nothing situation. Ironclad rules may be the only rules previously deceived investors and financial markets participants will interpret as rules in such countries. But this is only conjecture that so far, to our knowledge, has not been formally proved. In this sense, the debate surrounding the convenience and effectiveness of currency boards is perhaps a red herring that distracts from the real issues, which are how to determine the best policy rule for countries that have frequently reneged on commitments and how to protect those rules from the continuous assault of the time inconsistency forces. Economists and policymakers still have a lot of thinking to do on both counts, especially after the recent economic experiences of Argentina and Mexico.

Notes

I would like to thank Ken Emery, David Gould, and Owen Humpage for valuable comments on an earlier version of this article. I am grateful as well to Stephen Brown and Evan Koenig for insightful suggestions that were extremely useful in presenting my arguments. The editorial assistance of Rhonda Harris greatly improved the clarity and quality of the exposition. All remaining errors are mine.

[1] For a more detailed discussion of the tequila effect, see Zarazaga (1995a).

[2] Unfortunately, with the notable exception of Schwartz (1993), this insight has been lost in the currency boards literature.

[3] Although economists and social scientists have long been aware of this problem, (see, for example, Simons 1936), it was not until 1977 that it was formalized and brought to the forefront of the theory of economic policy by Kydland and Prescott (1977). For an excellent summary, see Taylor (1985).

[4] This article assumes the reader is familiar with the definition of the main monetary aggregates and, in particular, with the difference between primary expansion and secondary expansion of the money supply. See Zarazaga (1995a) for a brief and pedagogical exposition of these issues. For a more rigorous treatment, see Hanke and Schuler (1994) and Humpage and McIntire (1995).

[5] Law number 23,928.

[6] The stock of foreign reserves corresponds to the liquid foreign reserves net of domestic government dollar-denominated debt in the central bank's portfolio.

[7] The 2.3 billion pesos by which the monetary base exceeded the stock of foreign reserves at the end of March 1995 represented only about 1 percent of Argentina's GDP. It is unlikely that the demand for local currency will ever fall below that proportion of GDP, and, therefore, it was unlikely that in March 1995 Argentina's central bank would have had to buy back all the monetary base (12.3 billion pesos) with the $10 billion of reserves.

[8] This reference is to the 1994 insolvency of a small municipality in the United States that threatened to send that country's municipal bonds markets into a tailspin because of fear that other municipalities would default as well.

[9] For example, Hanke and Schuler (1994, 86) assert that "Failures by commercial banks have been minor in [currency board] systems." But the lessons that can be extracted from the historical experiences they reviewed are very limited because almost all such experiences have taken place in British colonies whose commercial banks were usually branches of international financial institutions. Those financial institutions had, as eloquently stated by Schwartz (1993, 182–83), "the resources to support a troubled local branch.…The London head offices of local branches provided lender of last resort services, if needed." In contrast, foreign banks were among the first to cut credit lines to their Argentine branches in the aftermath of the devaluation of the Mexican peso.

[10] It is important to emphasize that I do not claim that currency boards *create* banking crises, but rather that they have very limited ability to *prevent* them.

[11] This perception would not be totally unjustified. After all, as the next section explains, that is exactly what happened in Mexico at the end of 1994.

[12] According to Banco de México reports, on December 19, 1994, the stock of foreign reserves was $10.5 billion, while the monetary base was 59.6 billion new pesos. The dollar value of this monetary base at the exchange rate of 3.5 new pesos per dollar—that is, at the approximate exchange rate promised on the eve of the devaluation—implies a *MB$FR* ratio of 1.62.

[13] *MB$FR* ratios of 1.62 on December 19, 1994, and 1.12 on November 30, 1994, suggest explosive behavior in the intervening period. Indeed, in early December the monetary base grew about 22 percent, while foreign reserves fell around 16 percent. At least part of this expansion, however, may have been justified in the higher demand for currency typical of the month of December, when consumers need unusual amounts of cash to finance expenses related to Christmas.

[14] In fact, the depreciation of the peso was in that order of magnitude in the early phases of the floating exchange regime adopted after December 22, 1994.

[15] Although there is some debate about the underlying consequences of those capital outflows, it is symptomatic that foreign reserves fell by 40 percent in the twenty days immediately following a major political disturbance: the assassination of presidential candidate Luis Donaldo Colosio in March 1994. In fact, according to Calvo and Mendoza (1995), "Investors' prospects on Mexico's fundamentals *suddenly* changed, in part because of the increasing complexity of the ongoing political conflicts." [Emphasis added.]

[16] If Mexico's policymakers were mistaken in this regard, then they were in good company. As Calvo and Mendoza (1995) write, "Most of the information available until the end of 1994, including the assessment of international financial organizations, praised Mexico as a country with full balance in monetary and fiscal policies and set for strong future growth on the basis of its far-reaching reforms—*at about the same time the crash occurred, Mexico was accepted as a member of the OECD* [Organization for Economic Cooperation and Development]." [Emphasis added.]

[17] See, for example, Mancera (1995) for the Banco de México president's account.

[18] The inability to roll over the *tesobonos* debt (very short-term government debt adjusted according to the exchange rate) played a major role in the events that led to the crisis of December 1994. Interested readers can consult the study by Calvo and Mendoza (1995) and Cole and Kehoe (1995).

[19] Had Mexican monetary authorities had the recent econometric model of Kamin and Rogers (1995) and used it to predict the demand for currency, they would have forecast money demand growth below what they actually observed, especially for the first and third quarters of 1994. Had lower forecasts been used as a target in setting domestic credit (discount window) policies, the supply of monetary base would have grown at a slower rate than it actually did. Calvo and Mendoza (1995) use this finding to argue that monetary policy may have been too loose relative to the fixed exchange rate target and may have helped create the conditions for the speculative attack of late 1994.

[20] One could blame the policymakers for having missed several signs of the crisis to come. But many such signals could have been dismissed ex ante on the grounds that they reflected temporary factors containing very little information about more permanent economic imbalances. The exception, perhaps, is the money demand estimates mentioned in note 19. It is even possible to argue, as I do later, that Mexico was following a fixed exchange rate rule with an implicit escape clause, and that its policymakers merely exercised that escape clause in the face of extraordinary political events.

[21] Cottarelli (1993, Appendix II) points out that it is possible to identify countries—Belgium or Japan, for instance—whose central banks are not legally independent yet act much more so than the central banks of other countries that have theoretically independent central banks with the authority of written law. Cottarelli also discusses how the legal protection of the central bank can be and has been circumvented in the latter group of countries.

[22] See especially Lucas and Stokey (1983) and Chari and Kehoe (1990).

[23] This might explain why Canada, Belgium, and Italy have been able to sustain levels of government debt that, as percentages of GDP, are several times higher than the corresponding levels for Argentina, Brazil, and Mexico.

[24] It seems implausible that Mexico could restore its credibility with the simple announcement of a currency board law similar to Argentina's. Investors would question whether Mexico would adhere to yet another rule after abandoning its fixed exchange rate regime in October 1994.

[25] Examples of optimal rules are the Ramsey policies typically used as benchmarks of the analysis in the time inconsistency literature (see, for example, Chari 1988).

[26] Lucas and Stokey (1983), for example, construct models in which the optimal (Ramsey) policy is to abandon in the event of war the otherwise always honored rule of repaying the government debt. As Bordo and Kydland (1995) put it: "In an uncertain world, the Ramsey plan generally would be a contingent plan or rule. Strictly speaking, in a realistic environment the Ramsey plan would include many contingencies, some of which may make little difference to society's welfare."

[27] Bordo and Kydland identify these periods of suspension as 1797–1821 and 1914–25, which roughly correspond with the Napoleonic wars and World War I, respectively, and 1847, 1857, and 1866, which correspond to periods of banking panics.

[28] Bordo and Kydland (1995) state the problem well: "Drawbacks of including many contingencies, however, are lack of transparency and possible uncertainty among the public regarding the will to obey the original plan."

[29] Those who lean toward the second interpretation may point out that the assassination of presidential candidate Colosio qualified as a rare circumstance: no former or current president or presidential candidate has been assassinated in Mexico in the past fifty-six years.

[30] Thus, investors seem to have reacted not so much to fundamentals—that is, to economic policies—of the present but to those of the past. The same seems to be true about the causes of the bank panic that spread the tequila effect to Argentina, since according to a private report issued at the time, investors in that

country withdrew their money from the financial institutions on the concern that “the government might freeze bank deposits in order to stem a withdrawal of funds from the country” (according to a June 1, 1995, Bloomberg wire report) as it had done in 1990. The conjecture that Argentina’s and Mexico’s track records were catalysts of their financial crises could be of particular interest to scholars and policymakers because it suggests that reputation (and thus, past fundamentals) may play a major role in the genesis of herd behaviors like the one to which many analysts have attributed, at least in part, the speculative attacks against the currencies of Mexico and Argentina.

[31] This is an informal restatement of Chari’s (1988) advice that “policy recommendations that ignore the effect of history on people’s expectations will yield inferior outcomes” made in his insightful review of the extensions of the Barro–Gordon reputational framework to the case of contingent rules.

[32] Argentina’s people and policymakers also may have been inspired by the example of their close neighbor, Chile. That country’s rapid rate of growth over the past twelve years (GDP per capita has grown at an annual rate of almost 5 percent since 1983) is largely seen as the reward for the very strict monetary policies with which Chile responded in 1982 to a severe banking crisis. That crisis resulted in a decline of 15 percent in GDP and in unemployment rates in the same range as those Argentina is experiencing now.

References

Barro, Robert J., and David B. Gordon (1983), “Rules, Discretion, and Reputation in a Model of Monetary Policy,” *Journal of Monetary Economics* 12 (July): 101–21.

Bordo, M. D., and Finn E. Kydland (1995), “The Gold Standard as a Commitment Mechanism,” in T. Bayoumi, B. Eichengreen, and M. Taylor (eds.), *Modern Perspectives in the Gold Standard*, (Cambridge: Cambridge University Press).

Calvo, Guillermo, and Enrique Mendoza (1995), “Reflection on Mexico’s Balance-of-Payment Crisis: A Chronicle of a Death Foretold,” mimeo, Board of Governors of the Federal Reserve System.

Chari, V. V. (1988), “Time Consistency and Optimal Policy Design,” Federal Reserve Bank of Minneapolis *Quarterly Review*, Fall, 17–31.

———, and Patrick J. Kehoe (1990), “Sustainable Plans,” *Journal of Political Economy* 98 (August): 783–802.

Cole, Harold L., and Timothy J. Kehoe (1995), “Self-Fulfilling Debt Crisis,” unpublished manuscript, Federal Reserve Bank of Minneapolis and University of Minnesota.

Cottarelli, Carlo (1993), “Limiting Central Bank Credit to the Government: Theory and Practice,” Occasional Paper 110, International Monetary Fund.

Hanke, Steve H., and Kurt Schuler (1994), *Currency Boards for Developing Countries: A Handbook* (San Francisco: International Center for Economic Growth).

Humpage, Owen F., and Jean M. McIntire (1995), “An Introduction to Currency Boards,” Federal Reserve Bank of Cleveland *Economic Review*, Second Quarter, 2–11.

Kamin, Steven B., and John H. Rogers (1995), “Monetary Policy in the End-Game to Exchange Rate Based Stabilization: The Case of Mexico,” unpublished manuscript, Board of Governors of the Federal Reserve Board.

Kydland, Finn E., and Edward C. Prescott (1977), “Rules Rather than Discretion: The Inconsistency of Optimal Plans,” *Journal of Political Economy* 85 (June): 473–91.

Lucas, Robert E., Jr., and Nancy L. Stokey (1983), “Optimal Fiscal and Monetary Policy in an Economy Without Capital,” *Journal of Monetary Economics* 12 (July): 55–93.

Mancera, Miguel (1995), “Don’t Blame Monetary Policy,” *Wall Street Journal*, January 31, A20.

Schwartz, Anna J. (1993), “Currency Boards: Their Past, Present, and Possible Future Role,” *Carnegie–Rochester Conference Series on Public Policy*, 39 (December): 147–87.

Simons, Henry C. (1936), “Rules Versus Authorities in Monetary Policy,” *Journal of Political Economy* 44 (February): 1–30.

Taylor, Herb (1985), “Time Inconsistency: A Potential Problem for Policymakers,” Federal Reserve Bank of Philadelphia *Business Review*, March/April, 3–12.

Zarazaga, Carlos E. (1995a), “Beyond the Border: The Tequila Effect,” *Southwest Economy*, Federal Reserve Bank of Dallas, Issue 2, 7.

——— (1995b), “Can Currency Boards Prevent Devaluations and Financial Meltdowns?” Federal Reserve Bank of Dallas *Southwest Economy*, Issue 4, 6–9.

Should Bank Reserves Earn Interest?

Scott Freeman
Professor
University of Texas at Austin

Joseph H. Haslag
Senior Economist
Federal Reserve Bank of Dallas

Paying interest on reserves would increase the demand for deposits and thus for reserves. This, in turn, would raise the value of existing reserves, increasing the wealth of those who own bank deposits at the time that interest payments are initiated.

> The case for payments of interest on reserves applies not only to the 100% reserve system, but equally to our present fractional reserve system. Accordingly, even if reserves are not raised to 100%, Reserve Banks should be required to pay interest on their deposit liabilities.
>
> —Milton Friedman
> *A Program for Monetary Stability*

As the introductory quote indicates, Milton Friedman (1959), among others, has advocated paying interest on reserves.[1] In the United States and many other countries, banks and other financial intermediaries are required to hold a fraction of their assets as fiat money—unbacked, interest-free bills of the central bank. In the absence of interest on these reserves, the average return to assets held by banks must lie below the market rate of return. This implies that banks must pay their depositors a return below the market rate of interest, unnecessarily discouraging the holding of bank deposits. Because such intervention into the business of banking is so common, basic questions about the desirability of such requirements may easily be overlooked. For instance, by forcing banks to hold unbacked assets paying no interest, might the central bank be discouraging banking and the accumulation of capital?

But where would the interest come from? As with any government expenditure, interest paid on reserves must (at least eventually) come from taxes, raising two questions: Wouldn't wealth be reduced by the rise in taxes? Wouldn't taxation introduce its own economic distortions, possibly worse than those that result from the absence of interest?

Paying interest on reserves would increase the demand for deposits and thus for reserves. This, in turn, would raise the value of existing reserves, increasing the wealth of those who own bank deposits at the time that interest payments are initiated. Bruce Smith (1991) shows that this windfall gain to those holding deposits at the time the policy is enacted comes at the expense of future generations; that is, future generations must pay the taxes to finance the interest payments but do not receive all of the resulting benefits. Thus, Smith shows that the transfer of wealth created by the payment of interest makes future generations worse off.

In this article, we propose a means of eliminating this transfer. We begin by discussing the role reserve requirements play in a simple economy. People finance the next period's consumption by holding deposits. The key feature of the model is that reserve requirements force banks to hold fiat money as fractional backing for deposits. The merits of paying interest on reserves will be clear if the government offsets the wealth transfer identified by Smith.

Our idea for an offsetting transfer is adapted from a policy proposed by Leonardo Auernheimer (1974).[2] When interest on reserves is initiated, the central bank should expand the stock of nominal reserves to keep the price level from decreasing. If the central bank uses this increase in the money stock to purchase interest-bearing assets (an open market operation), the interest generated by these assets can help pay for the interest paid on reserves, lowering the tax burden on future generations. We argue that paying interest on reserves, when accompanied by the appropriate open market operation, can make every future generation better off without hurting initial deposit holders.

We also take up the second of our nettlesome questions: Would taxation introduce its own economic distortions? The taxes available to government in the real world are generally ad valorem taxes; the amount of tax collected is set at some fraction of an economic variable, such as income or sales. Ad valorem taxes artificially discourage the taxed activity, just as the absence of interest on reserves discourages deposits at banks. We show that despite this tax-induced distortion, we can make people unambiguously better off by paying interest on reserves. This improvement occurs even if the interest must be funded by a tax used in the real world—a distorting ad valorem tax on capital—if this tax is accompanied by a price-stabilizing open market purchase.

In sum, our questions about the costs of paying interest on reserves are fairly straightforward to resolve. Both capital taxation and open market operations are widely used real-world policy options. Therefore, there exists a way to finance the payment of interest on reserves that will make the public unambiguously better off.

A model of reserve requirement banking

To address these questions, let us examine a simple model adapted from the framework shared by David Romer (1985), Thomas Sargent and Neil Wallace (1985), Scott Freeman (1987), and Smith (1991) in which financial intermediaries that mobilize capital are subject to a reserve requirement.

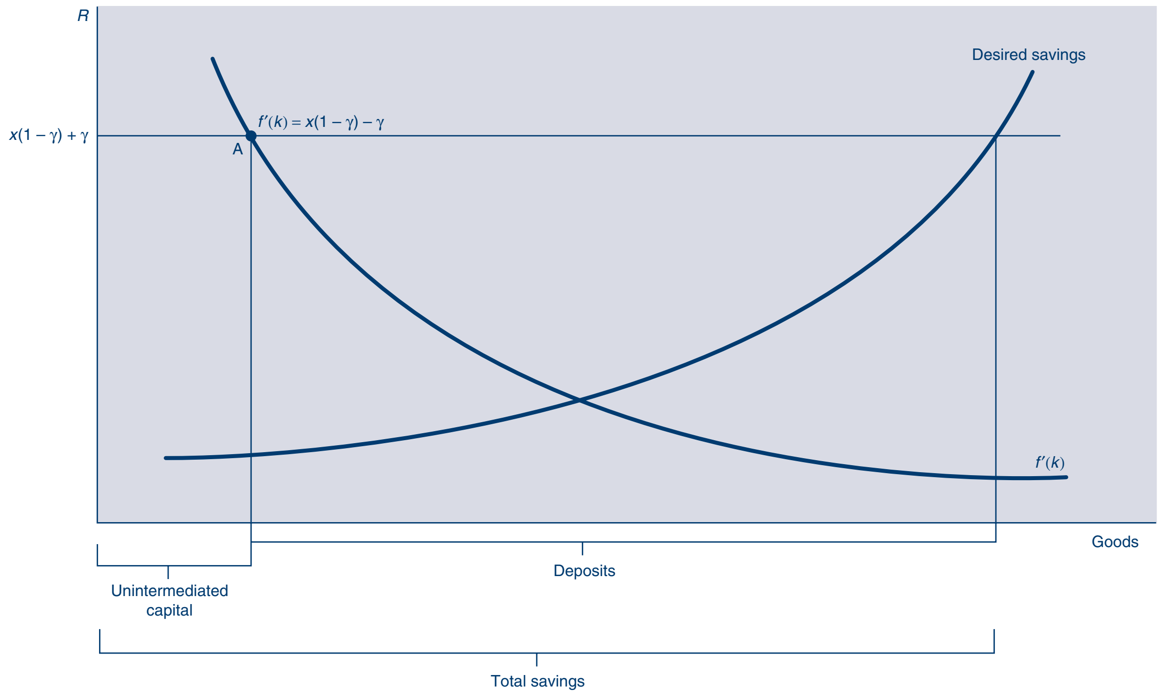
In each period, starting from some initial period 1, N people who live two periods are born. Each produces y goods when young and nothing when old, but wishes to consume in both periods of life. The problem facing these people is the means of financing consumption in the second period of life. There is also a generation that lives and consumes only in the initial period, hereafter referred to as the "initial old."

In the first period of this model economy, there is a fixed stock of M (divisible) pieces of paper called *fiat money*. In addition to money, there are also two forms of capital. The first form is available to any individual in isolation. An investment of k_t goods in period t will produce $f(k_t)$ consumption goods in period $t + 1$. The marginal product of capital, which we express as $f'(k)$, is positive but decreasing. The second form of capital produces a constant x consumption goods ($x > 1$) in period $t + 1$ for each good invested at t. This latter form of capital can be made only in amounts greater than y so that no individual alone has the resources to finance capital. Both forms of capital produce consumption goods only once.

Note that the second form of capital is illiquid in this economy because it cannot be divided into small units. It is easy to see how an intermediary can overcome this illiquidity by simply pooling the deposits of many individuals to an amount greater than y. We assume for simplicity that the intermediation services are costlessly and competitively provided by entities referred to as "banks."[3]

In this economy, we assume that a reserve requirement is imposed: for each good deposited, a bank must hold fiat money worth γ goods but is free to invest the remaining $1 - \gamma$ goods in the illiquid, or intermediated, capital good.[4] (We assume throughout this analysis that the initial old hold positive quantities of both unintermediated capital and deposits.) If fiat money's rate of return is less than that of capital, banks will hold no more than the required balances of fiat money. Suppose, for now, that banks do not hold any excess reserves. (We will verify shortly that this is a wise decision.) If s_t denotes deposits per young person, then banks will hold fiat money balances worth $\gamma N s_t$ goods. Those required reserves represent the total *demand* for fiat money measured in goods. The *supply* of fiat money is M dollars or, when measured in goods, $v_t M$, where v_t represents the goods that can be purchased by a single dollar. The goods value of a dollar is simply the inverse of the dollar price (p_t) of one good, or $v_t = 1/p_t$. Furthermore, the gross real rate of return from holding fiat money is the ratio of goods purchased by a single dollar

Figure 1
Determining Savings, Deposits, and Unintermediated Capital for a Given Reserve Requirement



in period $t + 1$ to the goods purchased by a single dollar in the current period, or v_{t+1}/v_t.

For the demand for fiat money to equal its supply,

$$(1) \qquad \gamma N s_t = v_t M.$$

Notice that when deposits, s_t, are constant over time, the demand for fiat money is constant over time. Therefore, when the stock of fiat money is also constant over time, the value of a dollar and the price level will both be constant over time. It follows that the gross real rate of return of a dollar, v_{t+1}/v_t, equals 1.[5]

What, then, will be the rate of return offered by competitive banks? Assuming for simplicity that intermediation services are costlessly provided by banks in a competitive market, then banks will offer depositors the rate of return that the banks can earn on the assets they hold. This (gross, real) rate of return (call it R) is

$$(2) \qquad R = (1 - \gamma)x + \gamma$$

because for each good deposited, the bank can invest $(1 - \gamma)$ in capital paying the rate of return x and purchase γ in fiat money paying the rate of return 1. Notice in equation 2 that increasing the reserve requirement lowers the rate of return on deposits by forcing banks to hold more low-return fiat money per deposit. Clearly, every subsequent generation suffers from this lower rate of return on their deposits. Moreover, with $x > 1$, equation 2 indicates that the bank best serves its depositors by not holding reserves in excess of those required.

People will invest in the asset paying the better rate of return. This implies that the people who hold both deposits and unintermediated capital will invest in unintermediated capital up to the point that its marginal rate of return just equals the rate of return offered by intermediaries:

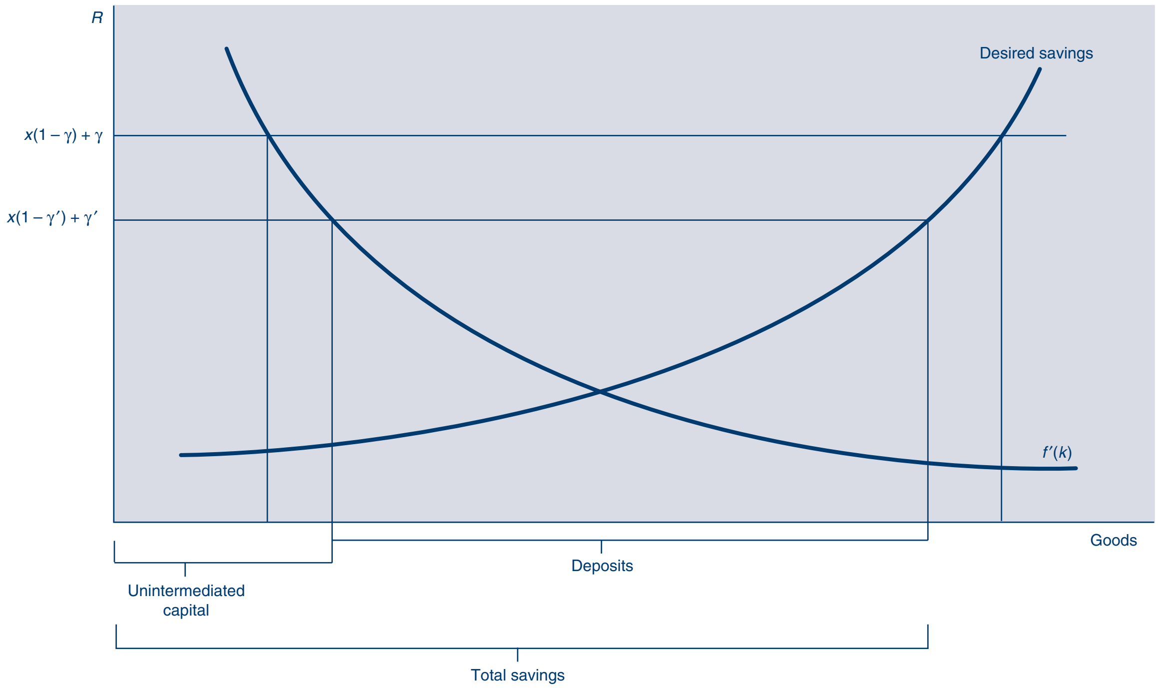
$$(3) \qquad f'(k) = (1 - \gamma)x + \gamma.$$

Because an increase in the reserve requirement lowers the return on intermediated capital, people switch from deposits to unintermediated capital. This switching occurs until the rate of return on unintermediated capital falls to equal the new lower rate of return on deposits.

Figure 1 illustrates the basic point made in equation 3. The desired savings curve plots the quantity of savings for next-period consumption at different rates of return. The rate of return on the vertical axis is equal to the rate offered by competitive banks and is determined by the returns on intermediated capital and reserves. The horizontal line emanating from the value $x(1 - \gamma) + \gamma$ on the vertical axis in Figure 1 is the return on deposits. For a given

Figure 2

The Effect of an Increase in the Reserve Requirement from γ to γ'



level of savings, the distribution between intermediated and unintermediated capital depends on the assumption that the return on unintermediated capital falls with each additional unit of this form of capital. Figure 1 captures this feature by representing the $f'(k)$ curve as a downward sloping line. From equation 3, people add units of unintermediated capital up to the point at which the return offered by banks equals the return on unintermediated capital. This occurs at point A in Figure 1. The horizontal distance between the vertical axis and point A measures how much unintermediated capital people will choose. The difference between desired savings and unintermediated capital—the horizontal distance between total goods saved and point A—measures the quantity of deposits.

An additional implication of equation 3 seen in Figure 1 is that with reserve requirements, the output produced by one more unit of unintermediated capital, $f'(k)$, is less than the output from a unit of intermediated capital, x.[6] Therefore, by encouraging people to switch from intermediated capital to unintermediated capital, a reserve requirement reduces output for each good switched. More generally, higher reserve requirements discourage total savings because of the lower rates of return offered on both unintermediated capital and deposits. Figure 2 illustrates the effects an increase in reserve requirements has on savings and each form of capital.

There is, we should note, a group that benefits from the imposition of a reserve requirement: the initial old. By assumption, this group starts with a portfolio of assets that include fiat money. If reserve requirements were removed, this fiat money would have no value. Consequently, the value of the initial old's portfolio would fall. Alternatively, increasing the reserve requirement increases the demand for fiat money, making each dollar more valuable and raising the welfare of the initial old by raising the value of fiat money (see equation 1). In short, the reserve requirement transfers wealth from all future generations to the initial old.

The central bank can increase the rate of return on deposits if it increases the rate of return on fiat money by, for example, paying interest on required reserves. This will increase the rate of return paid to depositors, but will it make them better off? As we demonstrate in the next section, the answer depends on how this higher rate of return is financed.

A case with interest payments on reserves

In this section, we consider how different financing schemes affect the desirability of paying interest on reserves. Paying interest on reserves will be deemed desirable if at least one group is made better off while no other group is harmed. In the model outlined above, the groups can be identified using the date at which the policy is implemented as the reference point; thus, the two groups that come to mind are those already holding money when the policy is implemented (the initial old) and the future generations.

The central bank as an intermediary. Consider first a policy that would have the government pay the interest from interest-bearing assets of the central bank. Suppose that instead of leaving the initial stock of central bank money in the hands of the initial old, the central bank takes it and uses it to purchase (intermediated) capital. This gives the central bank ownership of a stock of capital. (We focus our attention here on an equilibrium in which the stocks of reserves, central bank capital, and the value of money are constant over time.) Formally, the central bank's balance sheet constraint is

$$(4) \qquad K^g = vM,$$

where K^g is an interest-bearing asset that represents the value of capital held by the central bank.[7] The central bank will pay interest on its liability, reserves (central bank money), using the return on this capital net of its replacement cost, $xK^g - K^g = (x - 1)K^g$. If ρ denotes the nominal net interest paid on a dollar of reserves, then the interest paid on reserves equals ρvM, implying that each period the central bank's budget requires

$$(5) \qquad (x - 1)K^g = \rho vM.$$

Because the central bank owns capital exactly equal to the value of reserves ($K^g = vM$), the central bank can offer interest on reserves equal to the net return of capital:

$$(6) \qquad \rho = (x - 1),$$

which implies that the gross rate of return on reserves, $1 + \rho$, is x. Because the central bank backs its money with capital, reserves pay the same rate of return as other interest-bearing assets owned by private banks. Under this plan, the central bank has become an intermediary paying market interest rates to its depositors (private banks). Therefore, depositors at private banks will no longer care what fraction of their deposits is required to go into reserves. The gross rate of return on deposits is now

$$(7) \qquad R = (1 - \gamma)x + \gamma(1 + \rho) = (1 - \gamma)x + \gamma x = x,$$

which equals the gross rate of return on capital regardless of the size of the reserve requirement. For any positive reserve requirement, all future generations are made better off by this plan to pay interest on reserves because they are offered a higher rate of return on their deposits.

Would anyone oppose such a plan to pay interest on reserves from central bank capital? Yes, the initial old would. Notice that this financing scheme begins with the central bank confiscating the initial old's money balances without any compensation. Such a tax collection scheme reduces the wealth holdings of the initial old, reducing their consumption.

The payment of interest on reserves from central bank capital has the same welfare effects as abandoning reserve requirements. In both cases, future generations receive a better rate of return (x) on their deposits, but the initial old lose the value of their initial balances of fiat money.

There are three differences between abandoning reserve requirements and confiscating the initial old's money balances. First, when reserve requirements are simply abandoned, all fiat money becomes worthless.[8] Under central bank intermediation, however, there is still a demand for reserves, and the value of a dollar is again determined by the equality of supply and demand for reserves set forth in equation 1. As we have shown, banks will hold zero excess reserves, so that

$$(8) \qquad \gamma Ns = vM.$$

Second, the model economy described above specifies that intermediated capital always returns x units of the consumption good for every one unit invested. It is not difficult to imagine a situation in which returns are related to the quality of the investment decisions made. Under central bank intermediation, we entrust a governmental body, the central bank, with investment decisions. The central bank may not be motivated by maximization of profits. Consequently, if the central bank does not choose as wisely as private banks, the return offered on reserves may be below the market rate of return. Of course, one way to remove the investment decision from the purview of the central bank is to open the discount window. Banks could borrow funds at the market rate of return and make the investment decisions. Then the central bank's only responsibility would be to restrict its lending to sound banks.

Third, we have thus far assumed that intermediation services are costlessly provided. A more realistic assumption recognizes that costs, such as those of record-keeping, are associated with creating private intermediary services. With central bank intermediation, there is a second level of record-keeping; people make deposits at banks and then banks make deposits (hold reserves) at the central bank. If it is costly to keep records and otherwise manage deposits, the total of these costs will be higher under this two-level system of intermediation than under the one-level setup.

Tax-financed interest on reserves. Suppose that the central bank wants to finance interest on reserves without hurting the initial old. It can do so if each future generation is taxed to pay the interest.[9] Would the benefits of the increased rate of return exceed the cost of the taxation? Increasing the rate of return on deposits would increase deposits and thus capital, as desired. The increased deposits, however, also increase the demand for reserves. Greater demand for fiat money increases the value of the initial reserves owned by the initial old. (Note from equation 1, the equality of supply and demand for reserves implies that $v = \gamma Ns/M$. Clearly, an increase in s will increase v.) In effect, the taxes paid by future generations go to pay interest on reserves and to increase the wealth of the initial old. Smith (1991) demonstrates that this transfer of wealth from the future generations to the initial generation lowers the welfare of the future generations despite the greater rate of return on deposits. To understand this result, note that the taxes paid by people in the future generations are exactly equal to the value of the interest payments received on reserves. These two changes to lifetime wealth, therefore, exactly cancel each other out. However, the policy has a side effect: the reserves that the initial generation owns and that subsequent generations need have been made more expensive. This transfers wealth from subsequent generations to the initial old. Therefore, the central bank cannot increase the welfare of the future generations simply by financing interest on reserves through pay-as-you-go taxation.

Auernheimer (1974) suggests a way to finance the payment of interest on reserves without hurting or helping the initial old.[10] The initial old gain under the tax plan just described because of an increase in the value of their initial money balances. The value of money can be brought back to its initial level if the central bank prints more money, such that the increased demand for money is exactly matched by an increased supply of money. How can a plan featuring taxes and accommodating money supply increases help the future generations? Let the central bank use the increase in the stock of its money to buy capital. The central bank's exchange of (intermediated) capital for fiat money is an open market purchase. The additional capital can then be used to help finance the payment of interest on reserves, lessening the tax burden of future generations. Freeman and Haslag (forthcoming) show that this tax-financed interest on reserves makes the future generations better off. (A formal proof is also presented in the appendix.) The higher rate of return on deposits encourages savings through banks at its optimal level, without a transfer of wealth from the future generations to the initial owners of money. In short, the future generations pay enough taxes to finance the interest payments on reserves but do not pay for a transfer to the initial old.

A case with distortionary taxes

The financing scheme outlined above is based on a lump-sum tax. The desirability of taxing to pay interest on reserves may no longer hold if the tax, like many real-world taxes, itself distorts individual incentives. An income tax, for example, may well reduce incentives to work and invest, therefore causing more economic distortion than the absence of interest on reserves. To address this concern, we now examine the payment of interest on reserves financed by a tax commonly used in the real world, a tax on capital. We show that people are better off with interest paid on reserves, even if it must be financed by a tax that discourages the holding of capital.

Consider, in particular, a tax applied against the return from both types of capital; that is, the payment of interest on reserves is to be financed by a tax of α times the return to both intermediated and unintermediated capital. As with the lump-sum case described above, we assume that the government conducts an open market purchase that keeps the price level constant. Thus, the net interest on the government's capital goods plus revenue from the capital tax is equal to the government's net interest on reserves.

The question is whether interest-bearing required reserves are welfare-improving when financed with a distortionary capital tax. Freeman and Haslag (forthcoming) and the appendix to this article show that the total net return to the future generations is increased when the government pays interest on required reserves, even if the interest is financed by a tax on capital.

The intuition behind this result is fairly straightforward. If reserves pay no interest, a reserve requirement directly distorts the return to intermediated capital. In this way, the reserve requirement is like a tax on the return to intermediated capital, while unintermediated capital is not directly taxed. Paying the market rate of interest on reserves means that deposits earn the same return as unintermediated capital, ending the discouragement of deposits resulting from the lower return from required reserves. Taxing both intermediated and unintermediated capital at the same rate spreads the distortion equally, and thus efficiently, across the two types of capital. In short, people do not make investment choices between the two forms of capital based on after-tax returns. When taxes are applied equally, both the pre- and after-tax returns are equalized. The gain from the increased return on deposits more than offsets the lower after-tax return on unintermediated capital. Consequently, future generations have a higher total return than when the return of only one type of capital is distorted.[11]

The payment of interest on reserves encourages people to marginally substitute intermediated capital for unintermediated capital. For each extra unit of intermediated capital, x goods are produced, while an extra unit of unintermediated capital produces $f'(k)$ goods. We have seen that when intermediated capital is subject to reserve requirements without interest, $f'(k) = (1 - \gamma)x + \gamma = x - (x - 1)\gamma$, which is less than x. Therefore, when people switch one unit of savings from unintermediated capital to intermediated capital, more output is gained (x) from the increase in intermediated capital than is lost $[f'(k)]$ from the drop in unintermediated capital. Therefore, there is more overall output and greater welfare when interest is paid on reserves. Output and welfare would be even greater if the interest could be funded by lump-sum taxes, but stuck as we are with distorting taxes, the payment of interest on reserves is still an improvement.

Conclusions

In this article, we demonstrate how alternative schemes to finance interest payments on required reserves will affect people. We consider four different schemes: directly taxing initial required reserves, a lump-sum tax on future generations, and two financing schemes that are accompanied by open market purchases—lump-sum taxes and capital taxes.

We show that using lump-sum taxes accommodated by an open market purchase to finance interest-bearing reserves will avoid some of the pitfalls associated with either directly taxing initial required reserves or the lump-sum tax alone. When open market purchases accompany the payment of interest on reserves, members of the future generations are better off while the initial old are unaffected. Clearly, this makes society better off. We further show that paying interest on reserves is strictly better than not paying interest, even if the taxes are distortionary. This last result underscores the distortionary effect associated with reserve requirements. Spreading the distortion across both types of capital—in the spirit of the Ramsey rule of efficient taxation—raises welfare.

A key feature of the welfare improvement is the accommodating open market purchase suggested by Auernheimer. The payment of interest on reserves effects a transfer from future generations to the initial old. This transfer can be exactly offset by an open market purchase. The assets thus purchased can then be used to help finance the payment of interest. Such an accommodation is not beyond the central bank's normal operations. Indeed, Haslag and Hein (1995, 1989) provide evidence that the Federal Reserve systematically accommodates changes in reserve requirements with open market operations.

Overall, the main purpose of this article is to demonstrate that paying interest on reserves improves welfare in a broader class of model economies than previously believed. We extend the class of economies along two distinct lines. For some time, people have recognized the improvement that is possible in Friedman's setting with infinitely lived people and lump-sum taxes. Smith raises questions about the desirability of paying interest on reserves when the initial (finite-lived) money holders benefit but are not taxed. Our first extension shows that welfare improvement is still possible in this economy if a simple coordinated financing scheme is adopted. The second extension shows that paying interest on reserves can improve people's welfare, even if the interest is funded through distortionary taxes.

Notes

1 George Tolley (1957) also argues that the central bank should pay interest on reserves. Joshua Feinman (1993) traces the historical evolution of reserve requirements in the U.S. banking system. Feinman also notes that the Federal Reserve has explicitly supported legislation authorizing the payment of interest on reserves since the 1970s.

2 Auernheimer's proposal is designed to offset changes

in the demand for money induced by inflation rate changes. Also see Phillippe Bacchetta and Ramon Caminal (forthcoming), who apply the idea to reserve requirement changes.

[3] Certainly there are many other services provided by banks, but this one is simple to model and adequate to illustrate the points of this article. Other services of banks are implicitly included in *x*.

[4] In the United States, the requirement for checkable deposits at large banks is currently 10 percent, or $\gamma = 0.10$.

[5] More generally, if the economy is growing at the gross rate *n* (that is, $N_t = nN_{t-1}$) and the fiat money stock is growing at the gross rate *z* (that is, $M_t = zM_{t-1}$), the gross rate of return on a dollar will be n/z.

[6] From equation 3, $f'(k) = x - (x - 1)\gamma < x$.

[7] The central bank could also buy bonds from private banks, which would then use these funds to invest in intermediated capital. This scheme is closer to actual open market purchases but is equivalent in its effects to the direct purchases of capital by the central bank.

[8] This would not be true if there were an additional demand for fiat money as currency (negotiable notes passed from hand to hand). In most modern economies, the government retains a monopoly on the issuance of currency by outlawing its issuance by private banks backed by bank holdings of capital. This is exactly equivalent to a reserve requirement of 100 percent on currency.

[9] This is the financing scheme associated with Friedman's (1959) proposal and investigated by Smith (1991).

[10] Auernheimer (1974) describes just such a monetary policy accommodation scheme in describing the revenue-maximizing rate of inflation.

[11] The idea that taxing all goods improves welfare is discussed in Frank Ramsey's (1927) rule for efficient taxation. According to Ramsey, the government can raise welfare by setting distortionary taxes such that the percentage reduction in the quantity demanded of each commodity is the same. In our setting, Ramsey's rule is implemented by taxing both types of capital as opposed to taxing only one type. This result is demonstrated by Peter Diamond and James Mirrlees (1971) in a general setting. Diamond and Mirrlees demonstrate that taxing an intermediate input is not part of an optimal policy plan. In a monetary economy, Kent Kimbrough (1989) shows that the Ramsey tax rule applied to final goods improves welfare relative to a case in which intermediate goods were taxed.

References

Auernheimer, Leonardo (1974), "The Honest Government's Guide to the Revenue from the Creation of Money," *Journal of Political Economy* 82 (May/June): 598–606.

Bacchetta, Phillippe, and Ramon Caminal (forthcoming), "A Note on Reserve Requirements and Public Finance," *International Review of Economics and Finance*.

Diamond, Peter A., and James A. Mirrlees (1971), "Optimal Taxation and Public Production II: Tax Rules," *American Economic Review* 61 (June): 261–78.

Feinman, Joshua N. (1993), "Reserve Requirements: History, Current Practice, and Potential Reform," *Federal Reserve Bulletin*, June, 569–89.

Freeman, Scott (1987), "Reserve Requirements and Optimal Seignorage," *Journal of Monetary Economics* 19 (March): 307–14.

———, and Joseph H. Haslag (forthcoming), "On the Optimality of Interest-Bearing Reserves," *Economic Theory*.

Friedman, Milton (1959), *A Program for Monetary Stability* (New York: Fordham University Press).

Haslag, Joseph H., and Scott E. Hein (1995), "Does It Matter How Monetary Policy Is Implemented? *Journal of Monetary Economics* 27 (May): 311–26.

———, and ——— (1989), "Reserve Requirements, the Monetary Base, and Economic Activity," Federal Reserve Bank of Dallas *Economic Review*, March, 1–16.

Kimbrough, Kent (1989), "Optimal Taxation in a Monetary Economy with Financial Intermediaries," *Journal of Macroeconomics* 11 (Fall): 493–511.

Ramsey, Frank P. (1927), "A Contribution to the Theory of Taxation," *Economic Journal* 37 (March): 47–61.

Romer, David (1985), "Financial Intermediation, Reserve Requirements, and Inside Money," *Journal of Monetary Economics* 16 (September): 175–94.

Sargent, Thomas J., and Neil Wallace (1985), "Interest on Reserves," *Journal of Monetary Economics* 15 (May): 279–90.

Smith, Bruce, D. (1991), "Interest on Reserves and Sunspot Equilibria: Friedman's Proposal Reconsidered," *Review of Economic Studies* 58 (January): 93–105.

Tolley, George S. (1957), "Providing for the Growth of the Money Supply," *Journal of Political Economy* 65 (December): 477–84.

Appendix

In this appendix, we show more formally that paying interest on reserves will make people better off, even if the interest is financed with a distortionary tax on capital. To do so, we must first calculate the tax rate that would be needed to pay the market rate of interest on reserves. We let S represent total savings—deposits plus unintermediated capital—per young person and use asterisks to indicate values of variables in the absence of interest on reserves. The government must finance net interest on reserves of $(x-1)\gamma(S-k)$ from taxes on the return from savings, $\tau x(S-k) + \tau f(k)$, and from the interest on the capital it acquires from the open market purchase in the initial period, $(x-1)\gamma[(S-k)-(S^*-k^*)]$.

Altogether, this implies the government budget constraint is

$$(x-1)\gamma(S-k) = \tau x(S-k) + \tau f(k) + (x-1)\gamma[(S-k)-(S^*-k^*)], \quad \text{(A.1)}$$

or

$$(x-1)\gamma(S^*-k^*) = \tau x(S-k) + \tau f(k). \quad \text{(A.2)}$$

Paying interest on reserves makes future generations better off if for any given level of savings, $S = S^*$, the total return net of taxes is greater when interest is paid on reserves:

$$(1-\tau)x(S-k) + (1-\tau)f(k) > [x(1-\gamma)+\gamma](S-k^*) + f(k^*). \quad \text{(A.3)}$$

We can now use the government budget constraint (equation A.2) to cancel several of the tax terms with terms on the right-hand side of equation A.3, leaving us with

$$-xk + f(k) > -xk^* + f(k^*), \quad \text{(A.4)}$$

or

$$x(k^*-k) > f(k^*) - f(k). \quad \text{(A.5)}$$

We know that $k^* > k$ because unintermediated capital is taxed when interest is paid on reserves. Because $f(.)$ is a concave function (capital has a diminishing marginal product),

$$f'(k)(k^*-k) > f(k^*) - f(k). \quad \text{(A.6)}$$

When interest is paid on reserves, we know that the two forms of capital must offer the same marginal rate of return; that is, $f'(k) = x$. It follows that the inequality (equation A.5) is satisfied, proving that future generations are better off with interest paid on reserves, even if it must be financed through a distorting capital tax.