The Theory of Monetary Institutions

LAWRENCE H. WHITE
University of Georgia

BLACKWELL
Publishers

11
Monetary Rules

"Rules versus discretion for the central bank" has been the standard way of framing the debate over alternative monetary regimes since the 1930s, once it became clear that a discretionary regime was beginning to eclipse the rules of the gold standard. With the simultaneous rise of Keynesian economics, the pros and cons of discretion became identified with the pros and cons of "stabilizationist" or "counter-cyclical" monetary policy. Although the question of whether to have a central bank at all has been reopened since the mid-1970s (see Selgin and White (1994a), and chapters 12 and 13 in this book), proposals for fastening monetary policy rules onto the central bank continue to occupy center stage in debates over monetary reform.

Benefits and Burdens of Counter-cyclical Policy

In a "natural rate" economy, monetary policy is a potential source of deviations away from the natural rates of unemployment and output, conventionally shown as movements along the short-run Phillips curve (SRPC) and the short-run aggregate supply curve.[1] Such deviations are undesirable given that workers and producers want to make correctly informed decisions about job search and output, and prefer a less, to a more risky, macroeconomic environment. In such a world, what constructive role is there for monetary policy?

[1] The natural rates themselves can move for many reasons, such as shifts in the composition of the labor force, improvements in factor productivity, or raw material supply shocks.

The potential of counter-cyclical monetary policy does not lie in attempting to iron out *all* fluctuations of per capita real output around its historical trend. Monetary policy cannot usefully counteract swings due to technology or supply shocks that change the natural rate of output. The potential for monetary policy lies rather in avoiding the *component* of fluctuations that is attributable to monetary disequilibrium. In other words, the objective is not constancy of real output, but keeping the economy as close as possible to its natural rate of output. In a simple aggregate supply and demand framework, this means avoiding shifts in the aggregate demand curve, because aggregate demand shifts move the economy along the (upward sloping) short-run aggregate supply curve and, temporarily, off the long-run supply curve (which is vertical at the natural rate of output). When the economy is away from the natural rate of output, it is because agents are making misinformed decisions. The economy is regrettably discoordinated when real income is below, *or above*, its natural rate.

Viewing the task of monetary policy this way, and assuming that the sources of shifts in aggregate demand are variations in the quantity of money, M, or in the velocity of money V,[2] a successful counter-cyclical policy entails *offsetting changes in V with well-timed and correctly sized changes in M*. Activist monetary policy is a benefit on net if, and only if, it succeeds in this task.

Success is impossible if the economy rights itself faster than the monetary authority can ever respond to velocity shocks. Under strong-form rational expectations, the public anticipates any systematic monetary policy response to observable macroeconomic variables, and incorporates it into its pricing and output decisions, making monetary policy ineffective in stabilizing real income. Unanticipated policy can have a real effect, but its effect is not helpful: it only adds noise to the economy.

Success is not achieved in practice, even in conditions under which success is possible, if activist policy turns out to be cycle-amplifying, rather than cycle-dampening, because changes in money growth are poorly timed, or the wrong size. In the traditional monetarist diagnosis of typical central bank behavior, monetary policy moves are too often ill-timed, or ill-measured, because they act on the economy with a "long and variable lag." Forecasts of when a present change in monetary policy will begin to make an impact, and how far the economy will then be from its natural rate of output, are simply not good enough in the present state of knowledge. In too many cases, real output y has already returned, or nearly returned, to the

[2] When the "aggregate demand" curve is derived from the equation of exchange $MV = Py$ as the set of (P, y) pairs consistent with a given level of MV, then it is true, by construction, that the curve can only shift with a shift in M or a shift in V.

natural rate $y^\wedge$ by the time a positive boost arrives from higher money growth. The impact of policy is then to push y farther away from $y^\wedge$ rather than closer. To make things worse, the central bank tends, in practice, not to look ahead, i.e. to act in accordance even with the best forecasts available. Instead, it responds to political pressures to fight the current "number one evil" (Poole 1986).

Barro (1986) has pointed out another way in which discretion may be destabilizing: in contrast to the gold standard, discretionary fiat money regimes have contributed to real instability by "unanchoring" long-term price-level, and inflation rate, expectations. Changes in the expected inflation rate lead to changes in velocity, and thus create disturbances to aggregate demand.

In addition to arguments about the prospects for stabilization policy, a second strand of the traditional case for rules has come from concerns about the possible political (mal)incentives of monetary authorities. Public choice theorists identify discretion with the absence of a monetary constitution. Monetary authorities are free to pursue a political agenda, possibly seigniorage or political business cycles, contrary to the interests of the average citizen.

Since Kydland and Prescott (1977), the time-inconsistency problem has provided a third major strand to the case for rules. As we have seen, the Kydland–Prescott literature identifies discretion with the absence of a credible precommitment binding future monetary policy, leading to suboptimality in the form of excessive inflation. Unlike the traditional monetarist and public choice arguments, the suboptimality does not depend on the monetary authority's having too little information, or the wrong incentives.

Independence for the Central Bank

Before turning to specific monetary rules, we consider a distinct prescription addressed to some of the same concerns. Proposals for "an independent central bank" do not envision a monetary policy rule, but rather discretion vested in the hands of central bankers rather than elected officials. The basic motivation is to avoid the malincentive problem. Central bank officials are to be given greater insulation from control by elected officials, in the hope that this will better enable them to resist short-sighted demands for inflationary finance, election-year monetary stimulus, or artificially low interest rates. A non-partisan central bank, proponents hope, will pursue public-interest goals using scientifically favored techniques. The case for independence has been bolstered by comparative studies suggesting that countries with greater central-bank independence have experienced lower inflation rates.

While sharing the goal of low inflation, advocates of rules are some of the harshest critics of independence as a means. They argue that a central bank able to resist political demands is also able to resist public accountability for choosing the wrong goals, choosing the wrong techniques for attaining those goals, and using the techniques incompetently. Central bank officials are sometimes among the strongest advocates of independence. Critics fear that this is because the central bankers would find it comfortable to be answerable to no one.

Apart from whether it would be desirable, it is far from clear how much central bank independence from the legislative and executive branches is really possible. In the case of the Federal Reserve System, the President appoints its Governors. Congress created the agency, and can rewrite its mandate whenever it wishes, as it has, several times, over the years. To what extent can the Fed then afford to be unresponsive to pressures from Congress or the President?

Arguments for Rules

H. Geoffrey Brennan and James M. Buchanan (1981) define a "constitutional" monetary system as any regime that limits government's discretion regarding money, just as the First Amendment to the US Constitution limits the federal government's discretion regarding speech, press, and religion. Given such a broad definition, we then need to distinguish two very different sorts of constitutional regimes:

1 where a monetary authority is established with limited delegated powers prescribed in writing (this corresponds to the main body of the US Constitution), or otherwise generally understood, and binding, and
2 where government plays no monetary role, so that the provision of money is left to private enterprises bound by contract law (this corresponds to the First Amendment's injunction that "Congress shall make no law regarding . . .").

Both regimes impose limits on government, but very different sorts of limits.

Correspondingly, there are three basic schools of thought on the question of a monetary constitution.

1 The *discretionary central banking* school favors discretion or activism, and opposes the attempt to hem in the central bank with prescribed rules. In nineteenth-century Britain, the Banking School

opposed the limits on Bank of England note issue prescribed by Peel's Acts, though they favored the gold standard as a natural contractual constraint. In the twentieth century, the Keynesians have been the chief proponents of discretion.

2 The *constitutional central banking* school advises that the central bank should follow a specific formula. The nineteenth-century Currency School favored a 100 percent marginal reserve requirement on Bank of England notes. In the twentieth century, Monetarists, led by Milton Friedman, have offered much-discussed money supply formulas (considered below).

3 The "free banking" or *free market money* school favors decentralized and competitive money supply over central banking of either sort, and favors removing government from the monetary system. The nineteenth-century Free Banking School favored an end to Bank of England monopoly in London, and opposed the extension of Bank of England powers in the 1844 Act. In the twentieth century, the school was largely dormant until 1976 when F. A. Hayek's *Denationalisation of Money* (1990) was first published.

Before Friedman, Henry Simons (1936) had offered the classic case for rules. Simons made the preference for rules over discretion part of the "classical liberal" ideology, akin to the preference for "the rule of law" over arbitrary rule by authorities. Discretion creates uncertainty about, and subservience to, the whims of rulers. Simons declared that the ideal rule was to freeze M1, an ideal to which Friedman nearly returned. As a means to that end, Simons favored the "Chicago plan" banking reform which would make reserve requirements 100 percent. (Otherwise, M1 would vary as the currency-deposit ratio varies.) Freezing the money stock is a simple and clear rule, and would bring about a mild deflation as real income grows. Unfortunately, Simons noted, it does not accommodate changes in velocity that would cause the price level to vary. Its enforceability was in question because the 100 percent reserve requirement on demand deposits would encourage growth of near-monies. As a second-best short-term proposal, Simons favored a rule to stabilize the consumer price level.

Friedman's Proposals

Milton Friedman (1960, 1968) offered the "*k* percent rule" as part of a "framework for monetary stability." He doubted that the market by itself could provide a stable monetary framework, because he thought that US history showed fraud and overissue to be the typical outcome of free banking. Later,

in light of evidence to the contrary, he reconsidered this view (Friedman and Schwartz 1986), and came closer to a "free banking school" viewpoint. In Friedman's (1960) view, it was up to the government to control the money stock M, by preventing counterfeiting, fraud, and bank runs that would over-expand or over-contract M.

Friedman's principal objection to having the gold standard play these roles, rather than a system of rule-bound fiat money, was the resource costs of the gold standard (as discussed in chapter 2). He added that an international gold standard (or any fixed exchange rate regime) makes the domestic money stock subservient to the balance of payments. For a country with a large international trade sector, fixed exchange rates might be worth it, but for a country like the USA, he considered it undesirable to make the domestic sector (then 95 percent of GNP) adjust to shocks in the international sector (5 percent). Fiat money, and floating exchange rates, allow an independent national monetary policy, which can, in principle, be devoted to pursuing a more stable money growth path than a gold standard would produce. (In practice, it has not turned out that way.) Floating rates also eliminate the chief rationale for harmful exchange controls and trade quotas, that they are needed to safeguard the nation's reserves.

In his 1968 Presidential address to the American Economic Association, Friedman elaborated his view of the benefits of rules over discretion. Because real variables tend toward their "natural rates," monetary policy cannot control real variables; it can only disturb them in the short run. The real interest rate can be disturbed through the liquidity effect, but is independent of monetary policy in the long run. The unemployment rate can be disturbed by surprise inflation, but the long-run Phillips curve (LRPC) is vertical at the natural rate of unemployment (this was not yet a widely accepted idea in 1968). The aggregate supply curve of real output is, likewise, vertical at the natural rate of output. In such a "natural rate" world, monetary policy is ultimately limited to controlling some nominal variable, such as the nominal money stock M, the price level P, the level of nominal income Y, or the nominal exchange rate. Unanticipated policy can disturb real variables away from their natural levels, but it is best to avoid such disturbances. The proper goals for monetary policy are, therefore, to provide a stable nominal anchor, and to avoid being itself a source of disturbances. Monetary policy should not try to offset changes in real money demand where the central bank cannot be sure of doing more good than harm (in light of the problem of long and variable lags).

The harm-minimizing proposal Friedman offered in 1960 was the "k percent rule": make some monetary aggregate (either H, M1, or M2) grow at the rate of k percent per year, where k is constant, month in and month out. The choice of which M to target is to be decided by which has the most

stable velocity V, so that nominal income MV is relatively stable. Circa 1960, this criterion favored M2. The numerical value of k is to be chosen for its consistency with zero secular inflation. In terms of the dynamic equation of exchange,

$$gM + gV = gP + gy$$

Friedman's proposal involves solving (once and for all) for gM, having plugged in $gP = 0$ and appropriate long-run values for gV and gy. Looking back from 1960, Friedman found that gV was about -1 percent per year, and gy was about 3 percent. Together, these values indicated setting $gM = k$ at 4 percent per year.

Friedman rejected a price-level rule on the grounds that the link from ΔM to ΔP is too loose, the lags long and variable. An attempt to home in on P by trial and error may be destabilizing, i.e. involve over-shooting or endless oscillation.[3]

To supplement the k percent rule, Friedman offered measures to make M growth easier to control. Recalling the money-multiplier formula that $M = H(M/H)$, these measures were designed either to tighten the Federal Reserve's control over the monetary base H, or to reduce variability in the money multiplier M/H. In 1960, the vestiges of the gold standard remaining under the Bretton Woods system – the fact that foreign central banks could redeem dollars for gold – meant that the monetary base could be altered by foreign central bank redemptions. Friedman advocated severing this link between H and the gold stock (which was later done by President Nixon in 1971). The monetary base could also be altered at the initiative of domestic commercial banks, if the Fed felt compelled to honor their requests to borrow H from the Fed when the banks were otherwise unable to meet their reserve requirements. Friedman advocated eliminating discount-window lending of H, and instead imposing fines for reserve shortfalls. (This advice has not yet been adopted.)

To eliminate variability in M/H, Friedman, like Simons, suggested imposing 100 percent reserve requirements on all bank-issued components of the target aggregate. So that 100 percent reserves would not be onerous to banks, competitive interest is to be paid on commercial banks' reserve deposits on the Fed's books. Even if the 100 percent reserve requirement is not adopted, paying competitive interest on deposits at the Fed is advisable, because it would reduce the sensitivity of reserve ratios to market interest rates. (This proposal has not been adopted, presumably because

[3] For a specific price-level-stabilization rule that, its proponent argues, would not suffer from the over-shooting problem, see McCulloch (1991).

the Treasury and Fed would lose the income gained by the banks.) If 100 percent reserves are politically infeasible, Friedman advised at least making ratios uniform across all components of M2, so that shifts among accounts (e.g. from savings to checking) do not change the money multiplier. He also advised fixing the ratios permanently, so that the Fed, in its regulatory or revenue-gathering roles, does not interfere with its own monetary targeting.

Twenty-five years later, Friedman's (1987) prescription had evolved somewhat. His views on monetary theory and practice had not changed, and the goals remained the same: monetary policy should avoid being a source of disturbance, and should provide a stable nominal anchor. The specific proposals, however, had changed due to "public choice" considerations: a greater cynicism, if you like, nurtured by two-and-a-half decades of watching the Fed resist his and other proposals for monetary targeting. While Friedman still believed it would be desirable to stabilize the M whose V is empirically most stable, which points toward a relatively wide aggregate like M2, he noted that the Fed had been able to plead inability to hit M targets, and thus to avoid accountability. The best target for the sake of accountability is the narrowest: the monetary base. In light of the Fed's tendency to resist, or subvert, any restraint on its discretion, Friedman now viewed the generalized k percent rule as a "half-measure" because it leaves the Fed bureaucracy intact. With enforceability a leading concern, he now promoted a monetary base freeze as the "best real cure" for the instability of discretionary monetary policy.

Freezing the monetary base, H, eliminates the variability in money growth at the source. Moreover, it allows elimination of the Fed itself, hence banishes from the tent the "camel's nose" pushing for discretion. Without a positive growth path for any monetary aggregate to pursue, the Fed's Open Market Committee and bond traders could be released to seek employment elsewhere. The Fed's bank-regulatory and clearinghouse roles could also be eliminated, or transferred elsewhere. Reserve requirements, no longer needed for M targeting, could be phased out. Friedman suggested that commercial banks could again be allowed to issue currency. Though he did not say why this is desirable, it would buttress the H freeze, because it would allow banks to meet public shifts from deposits into currency without losing reserves of high-powered money (hence without contracting M). Thus, Friedman's later proposals moved him very close to, perhaps even into, the free-market money camp.

As Simons had, Friedman noted favorably that a base freeze would allow for a mild price deflation if real economic growth outruns innovations in the payment system that reduce demand for base money. For optimum-quantity-of-money reasons, a rising purchasing power of the dollar is beneficial to the base-money-holding public.

McCallum's Case for a Feedback Rule

Bennett McCallum (1989, pp. 336–51) has pointed out that activism (a regime in which the money growth rate responds to state of the economy) is not synonymous with discretion. Activism can be carried out according to a non-discretionary, pre-specified rule that holds at all times. It may then be possible to combine at least some of the potential stabilizing advantages of activism with the time-consistency advantages of rules. McCallum argues for a rule with feedback that would arguably avoid secular inflation more surely than a no-feedback k percent rule, and would also dampen price-level movements in the face of velocity and real income shocks. Given the lack of professional consensus on the macroeconomic "transmission mechanism," McCallum's objective is a modest feedback rule that "works" in the context of all the leading macro models: "a sensible monetary strategy would aim for a zero inflation rate on average and would not attempt to be highly ambitious with regard to its effect on cyclical variation of real variables. Most important . . . is the avoidance of abrupt changes in conditions due to monetary policy itself."

In Friedman's k percent rule for money growth, the value of k is set once for all time, with the hope that it will be consistent with zero inflation. Whether zero inflation actually obtains depends on whether the growth rate of velocity gV, and the growth rate of real income gy, turn out as expected. However, gV is hard to predict, in part because technical progress in the payments system occurs at seemingly random intervals. "Velocity drift" can drive the inflation rate away from zero. We have seen that Friedman assumed, by simple extrapolation of trend, gV of -1 percent and gy of 3 percent, and so recommended gM of 4 percent. McCallum notes that in fact, over the period 1954–1986, the realized value of gV was 2.5 percent per annum. Given realized gV was 2.5 percent, setting gM at 4 percent would have produced inflation gP of 3.5 percent, rather far from the zero inflation hoped for.[4]

McCallum proposes alternative rules with feedback to avoid the persistence of such prediction errors. Consider first a rule that, in contrast to the simple k percent rule, adjusts annual gM in response to changes in gV. The rule is formulated in terms of the M that the Fed directly controls, the mon-

[4] The rise in velocity, however, was not independent of the fact that the k percent rule was not followed. The main reason velocity rose over the period in question was that expected and actual inflation rose, and the main reason inflation rose was that money growth rose much higher than 4 percent per year. If money growth had been held to 4 percent per year, velocity would have risen less. The inflation rate may have missed zero, but it would have missed by less than 3.5 percentage points.

etary base, so the relevant V is the velocity of the monetary base. For this rule, we accept the assumption that gy is 3 percent, a value that McCallum notes has, in fact, obtained over almost all 20-year intervals (not counting World Wars or the Great Depression). Recalling again the dynamic equation of exchange

$$gM + gV = gP + gy$$

zero inflation ($gP = 0$) implies $gM + gV = 3\%$. The rule accordingly specifies

$$gM = 3\% - gV$$

where gV is the average over the previous four years (a period long enough to span the typical business cycle). Under this rule, nominal GNP will grow at 3 percent per year on average, even if velocity drifts, and inflation will be zero so long as real income growth gy is 3 percent.

McCallum also offers a slightly more complicated rule to deal with cyclical changes in gy. He reasons: "It seems likely . . . that cyclical fluctuations in real output and employment would be kept small if fluctuations in nominal GNP were minimized." Where Py denotes the natural log of nominal GNP and Py^* the "target" value of Py for the most recently observed period, the modified rule is

$$gM = 3\% - gV + .25(Py^* - Py)$$

Growth in the monetary base would be augmented when GNP is below path, and diminished when GNP is above path. The parameter value of .25 (which says that if nominal GNP is 1 percent below path, the central bank steps up money growth by one-fourth of one percentage point) was chosen to be small enough to avoid the problem of over-reaction that had concerned Friedman. McCallum reports that his simulation studies, nesting the policy rule in a variety of macro models, indicate robustly that following the rule would have yielded a more stable Py than the money growth path the Fed actually followed.

If McCallum's rule is a no-lose improvement over actual Fed policy, why has it not been adopted? The bureaucratic perspective, discussed in chapter 8, suggests that the Fed's officials will resist the imposition of rules because they value the prestige and importance that comes with discretion. There is no organized interest on behalf of imposing a rule. The subject of monetary policy rules is esoteric to the public, and to Congress. McCallum notes that post-Bretton Woods experience has not been so traumatic: although inflation did hit double digits, neither a Great Depression nor a hyperinflation

has occurred in most countries. Central bankers in many countries have exercised their discretion to bring inflation down from double digits to a range of 3–4 percent, and now speak of their resolve to maintain "price stability." Under those circumstances, the public is unlikely to agitate for a major institutional experiment.

Simple Versus Complicated Rules

Friedman's proposal for freezing the monetary base, and abolishing the Fed, chooses a radical solution to the enforcement problem. Any rule allowing the central bank to remain in business must be enforced against a real-life agency staffed by experts in the field who naturally prefer to have the discretion to use their expertise. To survive, a rule must be resistant to amendment by a legislature that might defer to the central bank's expertise.[5] To be effective,

1 the rule must explicitly prescribe the central bank's operating routine in terms of variables and actions that outsiders can readily monitor,
2 someone must actually do the monitoring to detect any central bank departures from the rule, innocent or not, and
3 some disciplinary mechanism must penalize departures from the rule.

For example, there might be automatic dismissal for officials if performance within a specified range is not achieved; or cash bonuses only if targets are hit. It is difficult to find historical precedent for such a system of operating rules, monitoring, and penalties or incentives being applied to any government agency in any nation.

A cynic will note that the central bank itself has an incentive to make monitoring more costly for its would-be monitors. It can try to rationalize apparent deviations from the rule as really only matters of incorrect measurement, distortion in the aggregate being measured, or an emergency (if the rule has an emergency escape hatch). Monitoring, either by Congress or the public, is more difficult the more complex the rule. Serious concern for the monitoring and enforcement problems therefore favors a monetary aggregate rule over a price-level rule (P is harder to measure unambiguously). Within the set of monetary rules, it favors a no-feedback rule over a feedback rule, zero growth over positive growth, and a monetary base rule over

[5] Timberlake (1985) examines the legislative history of the Depository Institutions Deregulation and Monetary Control Act of 1980, and finds that Congress accepted, at face value, the Federal Reserve's most dubious claims of a need for expanded Fed powers.

an M1 or M2 rule. The monetary base can be prescribed tightly because the central bank controls it directly; a broader aggregate must be allowed to vary within a broader band because the money multiplier varies outside of central bank control. The base can be measured unambiguously on the central bank's balance sheet, whereas what should be included in M1 or M2 is subject to change with market innovations.

Thus the strong suit of the monetary base freeze is that it is the most *enforceable* of rules. It is the only rule that really is like a constitutional prescription of what government shall *not* do: it prescribes that the government shall not expand the sum of its fiat money. The base freeze requires no agency to administer it, hence avoids the "camel's nose under the tent" problem of having in place an agency that has an inherent interest in lobbying for greater discretion. The durability of the rule (were it to be adopted) matters, because a more durable rule will deliver more of the benefits of precommitment. If a rule is not expected to survive for long, then it will not reduce uncertainty about long-term inflation. A durable rule, by contrast, will reduce inflation uncertainty (as we have noted the gold standard did) and, thereby, reduce the resource costs devoted to filling the demand for inflation hedges. If the enforceability issue is paramount, the logic of banning the camel's nose from the tent can be taken even further. If even the frozen authorized issue is a dangerous precedent, this would suggest that the most durable rule removes money from government's hands entirely.

1 How might rules that *prevent* the central bank from pursuing countercyclical monetary policy actually promote *greater* cyclical stability?
2 How does the time-inconsistency argument for rules, as articulated by Kydland and Prescott, differ from the traditional monetarist argument offered by Milton Friedman?
3 Would freezing the monetary base mean greater, or lesser, volatility of interest rates?
4 "Monetary policy can be quite effective in controlling the price level, but not in increasing the average level of output or, in practice, reducing the size of variations in real output. Monetary policy should therefore concentrate exclusively on the task of keeping the price level stable."
   (a) Explain the reasoning behind the first sentence.
   (b) Does that reasoning actually support a price-level rule, as suggested by the second sentence? Does it support any other policy prescriptions just as much?

12

Competitive Supply of Fiat-type Money

F. A. Hayek's much-discussed monograph on the "denationalization of money" (Hayek 1990) predicted that, in the absence of legal barriers, the market economy would deliver a stable system of competing private irredeemable currencies. Currency issuers would compete for customers by promising stable purchasing power in terms of some basket of commodities. A vigilant financial press would help to enforce competitive discipline such that issuers would find it worthwhile to uphold their promises. If a currency's value was not as promised, Hayek's argument went, it would lose so many customers that the issuer would want to correct the situation quickly.

A distinguishing characteristic of the Hayekian regime is that it relies on mere *promises*. Reliance on enforceable *contractual guarantees* of purchasing power would be tantamount to a regime of redeemability, and a commodity (or multi-commodity) standard.

A weakness of Hayek's discussion, in light of the general problem of "time-inconsistency" identified by Kydland and Prescott (1977), is its failure to show that the issuer will not want to break its promise of stable purchasing power. The profitability of staying in business may not outweigh the profitability of spending into circulation larger sums of money than are consistent with keeping the promise. Guillermo Calvo (1978) and Bart Taub (1985) show, indeed, that overissue in the extreme (a hyperinflationary burst) can be profit-maximizing for issuers of irredeemable or fiat-type money. In their models, the one-shot gain from hyperinflation exceeds the present value of the stream of returns from any sustained lower path of issues. Aware that a profit-maximizing issuer would want to hyperinflate, agents with rational expectations will not want to hold a fiat-type money unless the issuer can enforceably precommit to a specified path for the quantity of its issues at all future dates.

In light of these results, the feasibility of private fiat-type money is doubt-

ful. We cannot appeal to historical experience to reassure us that private issuers have solved the time-inconsistency problem with fiat-type money. All known private monies have been either full-bodied commodity monies (e.g. gold coins), or redeemable monies (e.g. gold-redeemable banknotes or dollar-redeemable deposits). Hayek's prediction that commodity money, and redemption contracts, would be dominated by fiat-type money in a competitive market is therefore not persuasive, without a theoretical resolution of the time-inconsistency problem facing private issuers of irredeemable money.

Even before Hayek wrote, Benjamin Klein had examined the feasibility, and efficiency, of the competitive supply of fiat money. Klein (1974, p. 424) addressed in particular "the possibility that firms may 'deceive' their customers by supplying more money than is anticipated." In summarizing his results, Klein claimed to have found feasibility even without contractual precommitment to a quantity path: "it is shown that if consumers and producers make the same estimate of the short-run profits from a policy of deception, then the equilibrium quantity of brand-name capital will insure that firms will not excessively overissue." Thus Klein's concept of "brand name capital" appeared to solve the time-inconsistency problem, at least in the context of Klein's model.

Taub (1985, p. 195), in reporting his contrary finding of non-feasibility, noted two differences between his and Klein's models. Klein had simply assumed a money demand function, and had imposed a particular form of non-rational expectations. Taub derived money demand from an overlapping-generations model, and imposed rational expectations.[1] It might be thought from Taub's discussion that Klein's feasibility result was perfectly valid, given Klein's assumptions. The question of which model, and which result, to "believe" would then simply depend on which assumptions were preferred.

A closer examination of Klein's model shows that, in fact, feasibility does not obtain, in the absence of perfect foresight. The concept of "brand name capital" does not solve the time-inconsistency problem in the case of fiat money. Even under Klein's assumptions, a private fiat money issuer would find it profit-maximizing to hyperinflate.

Klein's Model with Perfect Foresight

Klein proposes to examine the competitive supply of fiat money under perfect and imperfect foresight. Money is issued in distinct brands, with vari-

[1] John Bryant (1981), in a neglected paper, found that competitive provision of fiat money was either infeasible or inefficient. He also cited Klein, and implicitly contrasted Klein's model with his own overlapping-generations model.

ables for "brand j" money (issued by the jth firm in the industry) denoted by a subscript j. Under perfect foresight, all changes in the purchasing power of a currency are anticipated. The rising marginal cost of producing real balances of brand-j money (which here can be interpreted as the cost of endowing the money with greater transactions-facilitating properties) limits the profit-maximizing quantity of *real* balances produced, $(M/P)_j$. In profit-maximizing competitive equilibrium, where marginal cost equals price, the marginal cost of producing real balances equals the "rental price" obtained by the firm on its money. The equilibrium condition is

$$i_j - i_{M_j} = \mathrm{MC}_j \tag{12.1}$$

where

i_j = the nominal interest rate on bonds denominated in money j
i_{M_j} = the nominal interest rate paid on balances of brand-j money
MC_j = the marginal cost of producing real balances of brand-j money

The left-hand side of equation (12.1), the difference between the two interest rates, can be considered the "rental price" of brand-j money: it indicates the yield differential (opportunity cost) consumers are willing to bear to hold brand-j money balances rather than bonds. (This equation is Klein's equation [5], with simplified notation.)

In Klein's theory, real money balances are not generally assumed to be costless to produce, because money has to be endowed with the capacity to render transactions-facilitating services. The real resource cost to firm j of producing real balances is an increasing function of the quantity of real balances produced (marginal costs MC_j are rising) in the area of the firm's chosen output. In the limiting case, where the MC of creating real balances *is* zero, the competitive outcome implies $i = i_M$, which is the "optimal quantity" of money outcome discussed in chapter 5.

We have already seen an equilibrium condition very much like equation (12.1), namely the equi-marginal condition for a competitive issuer of redeemable deposits (see chapter 3),

$$i_{\mathrm{L}} - i_{\mathrm{D}} = C_{\mathrm{L}} + C_{\mathrm{D}} + Q_{\mathrm{D}}$$

The sum $(C_{\mathrm{L}} + C_{\mathrm{D}})$, the marginal cost of intermediating loans into deposits, is the equivalent of MC_j, the cost of "producing money balances." In the case of irredeemable money, the marginal liquidity cost of bank liabilities Q_{D} is zero.

The total *nominal* balances produced of brand-j money, M_j, and the price level measured in j-money units, P_j, are individually indeterminate, but they

are also of no consequence. The issuer's choice of the nominal unit in which to measure money j is akin to a soft-drink bottler's choice of whether to measure his output in liters or fluid ounces. Real price and quantity are independent of that decision.

The rate of monetary expansion gM_j and the inflation rate gP_j are likewise indeterminate and inconsequential *under perfect foresight*. For the public to hold money j, under perfect foresight and perfect competition, the potential impact of any anticipated inflation would have to be neutralized by an explicit interest yield i_{Mj} that fully compensates for future depreciation of money j's purchasing power. Rearranging equation (12.1),

$$i_{Mj} = i_j - \mathrm{MC}_j$$

Assuming the Fisher effect to hold at every moment, and recalling that, under perfect foresight, the anticipated inflation rate equals the actual inflation rate, the nominal interest rate on money-j-denominated bonds equals the real rate of interest r plus the actual (and anticipated) inflation rate

$$i_j = r + gP_j$$

Given that the quantity of real balances $(M/P)_j$ does not change, the inflation rate equals the nominal money growth rate

$$gP_j = gM_j$$

Thus, by substitution

$$i_{Mj} = r - \mathrm{MC}_j + gM_j \tag{12.2}$$

Equation (12.2) says that the explicit yield on money j must equal the real rate of interest minus the marginal cost of producing real money balances, plus full compensation for any (perfectly anticipated) dilution of the purchasing power of money j via money growth (Klein 1974, p. 427, n. 5). Essentially, any newly printed units of money j must be distributed to the holders of existing units, in proportion to their existing holdings, leaving holders indifferent to the printing up of new units. There is no profit to the firm from expanding the nominal money stock under this condition, and no loss either, given that real cost is a function only of real magnitudes (it is costless to add zeros to the currency).

This perfect-foresight analysis shows that the determinacy of real balances is what matters, and shows that hyperinflation is not the issuer's dominant strategy. It thereby counters the predictions of economists (Pesek, Friedman) who had argued that laissez-faire in fiat-type money must lead to

an infinite price level. Those authors had discussed open competition in fiat money production as though it meant that open counterfeiting were permitted. Without *distinguishable* brands of money, as Klein points out, competition would, of course, drive the quality of private fiat-type monies to zero. "Competition" without distinguishable brands would drive to zero the quality of *any* good whose quality is not detectable at the point of sale. (If all producers could counterfeit identical Coca-Cola cans, the profit-maximizing strategy would be, to them all, to put water, rather than cola, in Coca-Cola cans.) But that does not show infeasibility where distinguishable brands are permitted.

Time-inconsistency problems do not arise in Klein's perfect foresight case, because promise-breaking, or deception, is ruled out by assumption. A perfectly foreseen would-be overissuer would never have any customers. A firm cannot decide, in the future, to deviate from an announced policy, because perfect foresight, in effect, collapses the future into the present.

Klein's Model with "Imperfect Foresight"

To analyze potential problems of time-inconsistency or deception, Klein moves on to an imperfect foresight case of a particular sort. This section will reconstruct his analysis. The next section will criticize it.

Klein now assumes that consumers do not know gP_j perfectly, but must form an estimate gP_j^*.[2] Under perfect foresight, as we saw in equation (12.2), higher rates of nominal money growth, and inflation, must be offset by a higher explicit interest yield i_{M_j}. Under imperfect foresight, consumers must likewise be compensated for higher anticipated inflation gP_j^*. In addition, to forestall deception, an issuer can and must offset a greater degree of misbehavior – larger discrepancies between anticipated and actual money growth – by a higher explicit interest yield. In equilibrium, high-confidence (low-discrepancy) monies will command a premium in the form of a higher "rental price," $i_j - i_{M_j}$.

Competitive equilibrium on the supply side implies that a higher-confidence money, earning for its producer a higher rental, will be produced at a higher marginal cost. The marginal cost of producing real balances of money j now has two components, combined in optimal fashion: the cost of providing greater transactions services, and the cost of increasing confidence.

[2] Klein (1974, p. 437) states that "in equilibrium the prior probability expected rate of price change distribution will have a variance," but the stochastic structure is never specified. The variance of the inflation rate plays no explicit role in the model.

The present value of the firm's rental stream, attributable to the public's confidence in its money, is the firm's "brand name capital," denoted β_j. Embroidering on Klein's discussion, the size of j can be represented as follows. If the firm owns assets earning i_j, and issues money paying i_{M_j} as its liability, its real income net of interest payments, π/P_j, is

$$\frac{\pi}{P_j} = (i_j - i_{Mj}) \left(\frac{M}{P}\right)_j \tag{12.3}$$

To simplify, assume that the cost of providing services (but not the cost of generating confidence) is zero. In other words, assume that the cost of generating confidence is the only cost to finding people to hold the money. Then, the entire net income can be considered a stream of returns to the brand-name-capital asset β_j. Assuming β_j to be infinitely lived, and costlessly maintained, the present value of this infinite income stream is

$$\beta_j = \frac{(i_j - i_{Mj}) \left(\frac{M}{P}\right)_j}{r} \tag{12.4}$$

(This is Klein's equation [6].) That is, the value of β_j is the capital value on which the stream of real income $(i_j - i_{M_j})(M/P)_j$ represents a normal rate of return.

To interpret this result, consider two extreme cases.

1 If confidence were also costless to produce, then, in competitive equilibrium, the value of β_j would have to equal zero. Intuitively, as confidence becomes unlimited, the value of confidence capital goes to zero, because confidence ceases to be scarce.[3] In this case

$$i_j - i_{M_j} = 0$$

or

$$i_j = i_{M_j}$$

There is no difference between the rate of return on bonds, and the rate of return on money. The outcome is the "optimum quantity of money" produced competitively: because the marginal cost of pro-

[3] Klein (1974, p. 435) notes that "if confidence were completely costless to produce, the value of the jth firm's brand-name capital . . . would vanish." That is, β_j goes to zero as confidence becomes non-scarce. Earlier, Klein (p. 425) had incorrectly suggested that β_j goes to infinity.

ducing real balances (by intermediating bonds into money) is zero in all respects, the opportunity cost of holding money (rather than bonds) is driven to zero.[4]

2 If, for some reason, the real interest payments on money are zero, so that nominal interest on money just equals the inflation rate

$$i_{Mj} = gP_j$$

then recalling from the Fisher equation that, in equilibrium,

$$i_j = r + gP_j$$

we get by substitution into equation (12.4) above that

$$\beta_j = \frac{[r + gP_j - gP_j]\left(\frac{M}{P}\right)_j}{r}$$
$$= \left(\frac{M}{P}\right)_j$$

Firm j's brand-name capital exactly equals the real quantity of its money in circulation. The stock of j-money corresponds to net wealth for its issuer, because it is a zero-interest, zero-maintenance cost "liability" that finances the ownership of financial assets.

Is the Equilibrium Rate of Inflation Bounded under Imperfect Foresight?

With foresight imperfect, the rate of monetary expansion can be higher than the public expects. The money issuer can adopt a time-inconsistent policy, or practice what Klein calls "deception." The profitability of staying in business must now be compared to the profitability of unbounded money growth. If the costs of producing nominal money balances are zero, and the anticipated inflation rate is systematically below the contemporaneous rate of money growth, then the profit-maximizing rate of monetary expansion would be infinite.[5] Infinite expansion of money j at a moment in time would mean

[4] Wallace (1983) derives a similar result. Klein, like Wallace, entirely begs the question of how explicit interest could be conveniently paid on currency. On the consequences of a significant cost of delivering interest on currency, see White (1987) and White and Boudreaux (1998).

[5] For a simple illustration of this point, see the rays in figures 7.5 and 7.6, showing seigniorage for an issuer facing fixed inflation expectations.

a one-shot confiscation of wealth from anyone who accepts money j. If potential acceptors of j-brand money recognized this outcome, however, they would refuse to accept money j, and $(M/P)_j$ would never become positive.

Klein argues that the hyperinflationary outcome is not inevitable, because there may exist a stable equilibrium where the issuer's temptation to deceive is curbed by the profit stream available from non-deception. In re-examining Klein's model, however, we will find that such an equilibrium is not globally profit-maximizing.

Klein (1974, p. 436, eq. 7) incorporates imperfect foresight by revising the equation for the jth issuer's real profit flow (equation 12.3 above) in basically the following way. The anticipated rate of inflation of the jth money, gP_j^*, is assumed to be incorporated in the nominal interest rate i_j, but not in the (pre-announced) interest yield on the jth money, i_{Mj}. Then, holders of the jth money will demand $gP_j^*(M/P)_j$ in rebates, as compensation for anticipated inflation of gP_j^*. Again, abstracting from the costs of producing transaction services and confidence,[6] the issuer's real profit becomes

$$\frac{\pi}{P_j} = (i_j - i_{Mj})\left(\frac{M}{P}\right)_j + gM_j\left(\frac{M}{P}\right)_j - gP_j^*\left(\frac{M}{P}\right)_j \tag{12.5}$$

where

$\frac{\pi}{P_j}$ = real profit

$(i_j - i_{Mj})\left(\frac{M}{P}\right)_j$ = net real interest income

$gM_j\left(\frac{M}{P}\right)_j$ = gross real revenue from issuing new money, before rebates

$gP_j^*\left(\frac{M}{P}\right)_j$ = the portion of new money that must be rebated to holders of existing j-brand money in order to compensate them fully for anticipated inflation, i.e. in order to keep real demand and thus $(M/P)_j$ from shrinking.

To examine whether it is feasible that the profit-maximizing rate of monetary expansion is finite, we examine the implications of meeting the first-order (equi-marginal) conditions. The marginal profit of monetary expansion is:

[6] To incorporate these costs of producing real balances would mean adding a constant to equation (12.5). The marginal profit from nominal money growth (equation 12.6) would not be affected.

$$\frac{d\left(\frac{\pi}{P}\right)_j}{dgM_j} = \left(\frac{M}{P}\right)_j \left[d\left(\frac{i_j - iM_j}{dgM_j}\right) + 1 - \frac{dgP_j^*}{dgM_j}\right] \tag{12.6}$$

where

$\frac{d(i_j - i_{M_j})}{dgM_j}$ = is the fall in rental price resulting from unanticipated inflation; hereafter we denote this term by u

$\frac{dgP_j^*}{dgM_j}$ = the degree to which current inflation rate expectations adjust to current money growth, hereafter denoted v

Klein appears to regard dgP_j^*/dgM_j as a constant. Implicitly, then, he assumes that inflation-rate expectations are determined by an equation of the form

$$gP_j^* = u + vgM_j \tag{12.7}$$

Perfect foresight is represented by $v = 1$; imperfect foresight by $v < 1$.

Klein (1974, p. 436) notes that if the issuing firm can hold $(i_j - i_{M_j})$ constant as gM varies, so that $u = 0$, and, if expectations adjust less than fully, so that $v < 1$, then the marginal profit of monetary expansion "is always positive, and therefore the firm can make its current profit rate as large as it wants by merely making gM_j arbitrarily large . . . The profit-maximizing rate of increase of money is therefore infinite." [7] Using our notation, if $u = 0$ and $v < 1$, then the marginal profit from faster monetary expansion is always positive ($u + 1 - v > 0$), and the firm is driven to inflate without limit. The issuer need not rebate all newly issued money to existing moneyholders, but can keep and spend a share of it, $(1 - v)gM$. A higher rate of monetary expansion gM is then always more profitable, because the value of the non-rebated share $(1 - v)gM$ is larger.

Klein (1974, pp. 436–7) then denies that this outcome will actually obtain:

[7] Klein (1974, p. 436) refers to the case of $v < 1$ as a case where "there are lags in the adjustment of anticipations," but talk of "lags" is not appropriate in a one-period model. The adjustment of expectations is partial, but it is contemporaneous and not lagged, as Klein (p. 437 n. 17) elsewhere notes. For the same reason, it is not strictly correct for Taub (1985, p. 195) to speak of Klein's using "adaptive expectations." The expectations in question are not even really forward-looking. The distinction between Klein's two cases is not really perfect versus imperfect *foresight*; it is more a matter of perfect versus (supposedly) imperfect *perception* of *contemporaneous* money growth. We retain the terminology of foresight and anticipations, for convenience.

> However, this argument assumes that the money firm's brand-name capital is constant and so fails to consider the effect on consumer confidence and the firm's demand from the policy of "deceiving" customers. . . . The higher the actual rate [of monetary expansion] compared to the anticipated rate . . . the lower will be consumer confidence. As β_j falls . . . [$i_j - i_{Mj}$] must also fall to keep [demand] constant. . . . Consumers can (and will) control [$d(i_j - i_{Mj})/dgM_j$] to prevent an infinite rate of growth of money.

He concludes (p. 438) that, as long as consumers do not underestimate the short-run gain from deception, and make u too small in absolute value, "wealth-maximizing firms will not inflate at an infinite rate."

Klein argues, in other words, that hyperinflation will not be profit-maximizing, once we take into account the fact that money growth reduces consumer confidence and, thus, reduces the rental price firm j can earn on its money, i.e. once we drop the assumption that u is zero.

For maximum profit to occur at a finite monetary expansion rate, we need

$$0 = (u + 1 - v)\left(\frac{M}{P}\right)_j$$

which implies

$$u = v - 1$$

and if $v < 1$, this requires $u < 0$.

Klein provides the first-order condition for maximum profit at a finite inflation rate. He does not, however, inquire whether a local maximum, at which the first-order conditions are met, is also a global maximum. We return to the jth firm's profit function (equation 12.5). Because $[gM_j - gP_j^*]$ is directly proportional to gM_j, and the firm acts to keep $(M/P)_j$ constant, the product

$$(gM_j - gP_j^*)\left(\frac{M}{P}\right)_j$$

(the sum of the second and third RHS terms in equation 12.5) grows without limit as gM_j grows. To keep π / P_j from also growing without limit, the firm's rental stream (the first term on the RHS of equation 12.5) must become negative without limit. This requires that the rental price $(i_j - i_{Mj})$ become negative, and negative without limit as gM_j grows without limit. If $(i_j - i_{Mj})$ is bounded below by zero, this condition cannot be satisfied. Maximum profit does not occur at a finite monetary expansion rate, but at an infinite rate.

It is fairly straightforward to explain why the rental price $(i_j - i_{M_j})$ might be bounded below by zero. If the rental price becomes negative, then $i_{M_j} > i_j$. The yield on the jth money would exceed the yield on bonds denominated in the jth money. In this event, bond holders would entirely abandon bonds for money, as the yield on money overtook the yield on bonds. Only the money issuer would be left to hold j-denominated bonds.

If $(i_j - i_{M_j})$ does not fall below zero, the change in the rental stream cannot continue, indefinitely, to offset increasingly large money-printing revenue. Klein notes, in an aside, that "an infinite inflation rate [would be] implied . . . if the absolute value of $[\mathrm{d}(i_j - i_{M_j})/\mathrm{d}gM_j]$ never reached $[1 - \mathrm{d}gP_j^*/\mathrm{d}gM_j]$." The issue, however, is not whether this plateau $(-u = 1 - v)$ is ever reached; it is whether it can continue be occupied throughout the relevant range.

The implications of $(i_j - i_{M_j})$ being bounded below by zero are shown graphically in figure 12.1. Following equation (12.5), the real profit π/P_j at various rates of monetary expansion is the sum of two terms. The first term,

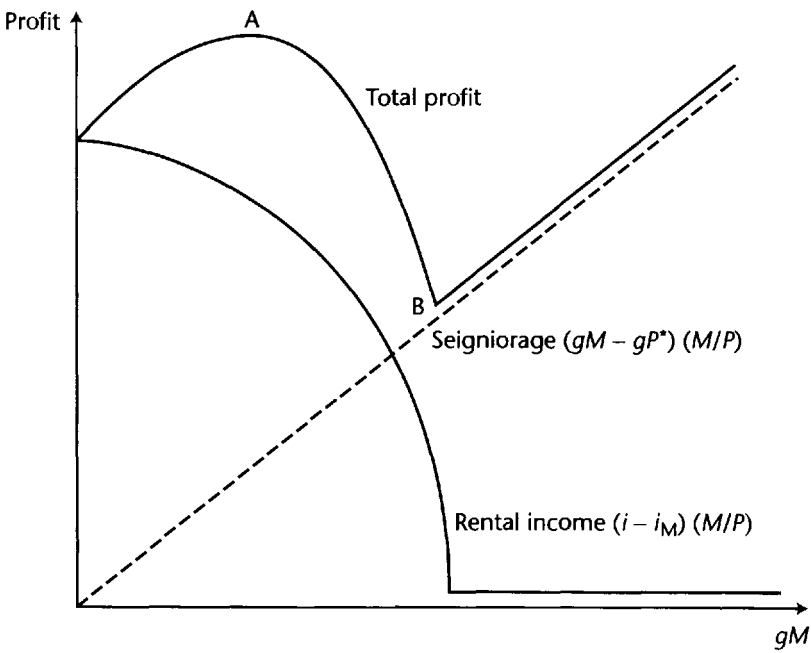

Figure 12.1 Unbounded monetary expansion: hyperinflation maximizes the issuer's profit when the public has imperfect foresight and the rental price on money is bounded below by zero

$$(i_j - i_{M_j}) \left(\frac{M}{P}\right)_j$$

is represented by the negatively sloped curve. The second term

$$(gM_j - gP^*_j) \left(\frac{M}{P}\right)_j$$

is represented by the dashed ray, which comes from the origin under the simplifying assumption that the constant in equation (12.7) is zero. The slopes of the curves representing these terms are respectively u and $(1 - v)$. Klein never specifies a function relating the first term to the rate of monetary expansion. We have drawn the curve so that there is indeed a local equilibrium at point A, where $-u = 1 - v$. As the first term asymptotically approaches zero, u approaches 0, so that beyond point B, $-u$ is less than $1 - v$, and profit increases with the rate of monetary expansion. Beyond point B, the issuer can effectively travel out a ray, just as if there were absolutely fixed inflation-rate expectations. Profits are globally maximized with infinite monetary expansion and infinite inflation.

Klein argues that u will always be sufficiently negative, because "consumers will . . . trade off higher levels of β_j, with correspondingly higher costs of holding cash balances $[i_j - i_{M_j}]$, against higher levels of unanticipated $[gM_j]$." As the jth firm (conjecturally) raises its money growth rate, it will forgo a higher rental price on its money. Consumers stand ready to pay a higher rental price for a money with a lower growth rate, precisely because they understand that, otherwise, the issuer would find hyperinflation profit-maximizing. Solving the problem of cheating, in this way, amounts to re-introducing perfect foresight through the back door. The public underanticipates money growth gM_j, but it knows exactly at each moment to what degree it is doing so (and by exactly how much it needs to lower the rental price it is prepared to pay as unanticipated money growth rises, in order to keep the issuer's profit from rising). To know exactly the discrepancy between actual and anticipated money growth is to know actual money growth; it is to have perfect foresight.

Conclusion

That a profit-maximizing private issuer of inconvertible money would hyperinflate means that the time-inconsistency problem bedevils private fiat-type money production even in Klein's model. The presence of "brand name capital" does not solve the problem.

Two solutions to the time-inconsistency problem with regard to money

issue are available, but both entail a monetary regime unlike Klein's or Hayek's.

1 As Taub indicates, time-inconsistency could be eliminated, even with irredeemable currency, if it were feasible to write, and enforce, a contract stipulating the future quantity of money to be issued from now to eternity. The feasibility, and enforceability, of such a contract is doubtful, however.
2 The traditional approach to binding a private money issuer is to write a contract obligating the issuer to buy back his money at a pre-determined price, i.e. a redemption contract. At least for money, redemption contracts would appear to be cheap to write and enforce.

Both kinds of contracts are seen in non-monetary settings, for example where artists sell lithographs or firms sell "collectors' items." A producer, who is selling a good above its marginal cost of physical production, wants to make it credible that he will not later drive the resale value down by selling more at a lower price (Coase 1972). Purchasers of a lithograph, typically, prefer a quantity guarantee (the promise of a limited number of copies), accepting the risk of a decline in resale value in order to enjoy the potential for the lithograph to appreciate. Holders of a medium of exchange, by contrast, would understandably prefer a value guarantee.

Questions

1 Under what conditions would private issue of fiat-type money reduce the opportunity cost of holding money (rather than bonds) to zero?
2 The time-inconsistency problem leads the fiat-money-issuing central bank in the Barro–Gordon model to a positive, but finite, inflation rate. Time-inconsistency leads the private issuer of fiat-type money in the Klein model to hyperinflation.
   (a) What accounts for this contrast?
   (b) Does the contrast indicate that a monopoly issue of fiat-type money is generally more trustworthy than competitive issue?