Inflation Targets and Inflation Targeting

Laurence H. Meyer

There is widespread agreement that price stability (in practice, low and stable inflation) should be an objective of monetary policy. This agreement is reflected both in the mandates set for monetary policy by governments and in the practice of central banks. Several other important questions about the objectives for monetary policy are less settled: Should there be other objectives? If there are multiple objectives, should one of the objectives take priority? And how explicit should the objectives be?

Central banks typically operate under one of two types of mandate. A hierarchical mandate makes price stability the primary objective for monetary policy and subordinates other potential objectives. A dual mandate recognizes two objectives—price stability and full employment—and puts them on an equal footing. Either regime could make the price stability objective more precise by setting an explicit numerical target for inflation.

Thus we can describe a typical central bank's mandate and objectives in terms of two sets of alternatives: between a hierarchical or a dual mandate, on the one hand, and an implicit or explicit inflation objective, on the other hand. During the 1990s, a number of central banks adopted a framework that is called inflation targeting, combining a hierarchical mandate and an explicit inflation objective. The United States, in contrast, combines a dual mandate and an implicit inflation objective.

Most of the discussion in the United States on the subject of mandates and objectives has been about whether to identify inflation as the single or primary objective and whether to move to a formal inflation-targeting regime.[1] The title of my lecture—"Inflation Targets and Inflation Targeting"—is intended to differentiate between two options for changing the policy mandate for the Federal Reserve. One option, which I favor, is setting an explicit numerical target for inflation within the context of our current dual mandate. The other option, which I do not favor, is moving to an inflation-targeting regime—that is, also substituting a hierarchical mandate for our current dual mandate. The purpose of this lecture is therefore to explain the benefits of an explicit inflation target in the context of the Federal Reserve's dual mandate and to set out the operational steps for implementing such a target.

Before proceeding, let me note that the views that I am presenting here are my own. I am not speaking for the Board of Governors or the Federal Open Market Committee (FOMC).

THE EVOLUTION OF MONETARY POLICY MANDATES

A good starting point is a survey of mandates around the world. I will begin by discussing the evolution of the mandate in the United States, including the precise language related to the dual mandate, the way in which the price stability objective has been interpreted, and proposed legislation that would have amended the mandate. Then I sketch an inflation-targeting regime and discuss some common elements and differences among the inflation-targeting regimes of developed economies around the world.

The Evolution of Policy Objectives in the United States

In the United States, it took quite some time for the Congress to establish a precise set of objectives for monetary policy. In fact, remarkably little about policy objectives was included in the original Federal Reserve Act in 1913. The only policy objectives of the Fed, as identified in that statute, were "to furnish an elastic currency [and] to afford means of rediscounting commercial paper." The absence of any mention of price stability undoubtedly reflected confidence that the gold standard, under which the United States was operating, would promote price stability. The intent of providing an elastic currency and of rediscounting commercial paper was to expand the supply of money and credit to accommodate expansions in production and the accompanying demand for credit. Given that the immediate impetus of the founding of the Federal Reserve was the Panic of 1907, promoting financial stability was a clear focus. The framers' intention was that the Federal Reserve would provide banks with a source of liquidity through rediscounting to meet deposit withdrawals.

Laurence H. Meyer is a member of the Board of Governors of the Federal Reserve System. These remarks were originally presented at the University of California at San Diego Economics Roundtable, San Diego, California, July 17, 2001.

[1] See, for example, Bernanke et al. (1999) and Gramlich (2000).

On several occasions during the 1920s and 1930s, the Congress debated a price-stability objective for the Fed. The Fed opposed such a mandate and it was not adopted. Congress did take a step toward a more explicit treatment of policy objectives in the Employment Act of 1946. This act identified the objectives for the government in general, but not specifically for the Fed. Still, the act was generally viewed as applying to the Fed, as a part of government. The objectives identified in the act were "to promote maximum employment, production, and purchasing power." Although this set of objectives could be interpreted as including both full employment and price stability, the title of the bill and the specific language suggests that the priority at the time was more to maintain full employment than to promote price stability. Such a focus on stabilizing employment and a relative inattention to inflation was perhaps an understandable reaction to the Great Depression when, for a decade, high unemployment and falling prices were the major problems facing the U.S. economy.

The specific mandate for the Federal Reserve was first added to the Federal Reserve Act in 1977, although that same language had been included in a 1975 concurrent resolution of the Congress. The 1977 amendment required the Board of Governors and the FOMC to "maintain the growth of monetary and credit aggregates commensurate with the economy's long-run potential to increase production, so as to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates." This language makes the objective of price stability explicit. Because the Fed can contribute to moderate long-term interest rates principally by achieving low and stable inflation, that objective is generally not viewed as an independent one. In addition, the goal of maximum employment is usually interpreted as maximum sustainable employment—meaning the highest level of employment that can be maintained without upward pressure on inflation. The mandate is therefore interpreted as a dual mandate: full employment and price stability.

The Federal Reserve has not set an explicit, numerical objective for inflation. Paul Volcker offered the following definition of price stability in 1983:

> A workable definition of reasonable "price stability" would seem to me to be a situation in which expectations of generally rising (or falling) prices over a considerable period are not a pervasive influence on economic and financial behavior. Stated more positively, "stability" would imply that decision making should be able to proceed on the basis that "real" and "nominal" values are substantially the same over the planning horizon—and that planning horizons should be suitably long.[2]

Alan Greenspan has described the price stability objective in a similar way: "We will be at price stability when households and businesses need not factor expectations of changes in the average level of prices into their decisions."[3]

These definitions make clear a commitment to low inflation. But they leave open whether, for example, the inflation rate prevailing today—about 2.5 percent for the core consumer price index (CPI) measure of consumer prices—is consistent with this definition. Is policy going to be set to lower inflation over time, and if so, by how much? These definitions also leave open the possibility of changing interpretations as the FOMC membership changes over time.

The Fed often prefers to state its objective without specifically mentioning price stability. This is perhaps because the emphasis on price stability is taken by some as carrying a hint of restrictive policy and as an inclination to always be leaning against cyclical increases in demand. The Fed sometimes prefers to state its objective simply as promoting maximum sustainable growth. Stating its objective in this way allows the Fed to offer a more positive message and leaves implicit the price stability objective in two ways. First, if the economy were to grow above a sustainable rate for long enough, overheating and higher inflation would eventually follow. Second, price stability contributes to a high and perhaps faster rate of growth in productive capacity, a point I will return to below. Nevertheless, I prefer to state the Fed's objectives as full employment and price stability. In my view, the Fed has no growth objective. At full employment, the rate of growth will automatically be the maximum sustainable rate the economy is capable of achieving and a rate largely independent of monetary policy, except insofar as monetary policy is successful in achieving price stability.

In recent years, bills have been introduced on a few occasions that would have made price stability

[2] Volcker (1983).

[3] Greenspan (1994).

the sole or primary objective for monetary policy and required the Fed to set an explicit numerical inflation target. In 1989, 1991, and 1993, Representative Steve Neal, Chairman of the House Banking Committee's Subcommittee on Domestic Monetary Policy, introduced resolutions instructing the Federal Reserve "to adopt and pursue monetary policies leading to, and then maintaining, zero inflation." In the 1991 and 1993 versions, zero inflation was defined as "when the expected rate of change of the general level of prices ceases to be a factor in individual and business decisionmaking." While these resolutions did not pass, the definition of price stability in the 1991 and 1993 resolutions was, undoubtedly not by accident, nearly identical to the language used by Chairman Greenspan and to the concept articulated earlier by Chairman Volcker.

A second set of bills was introduced by Senator Connie Mack and Representative Jim Saxton in 1995 and 1997. These bills instructed the Fed to set an explicit numerical definition of price stability and to "maintain a monetary policy that effectively promotes long-term price stability." Representative Saxton introduced a significantly revised version of these bills in 1997 and 1999, mandating price stability as the "primary goal" of the Federal Reserve and requiring the Fed to establish an explicit numerical definition of inflation. Senator Mack reintroduced his version in 1999.

I interpret these bills as attempts to push the United States toward a full inflation-targeting regime. Indeed, the Mack versions would establish an inflation-targeting regime among the strictest in the world, given that it would have established price stability as the sole objective of monetary policy, not simply a hierarchical set of objectives. The Saxton version is more in line with hierarchical mandates employed in many inflation-targeting regimes. These bills were, therefore, vigorously opposed by advocates of the dual mandate. Perhaps because these bills formed the backdrop to the debate in the United States about the policy mandate, little discussion has taken place on the merits of moving to an explicit numerical inflation target in the context of the prevailing dual mandate. Of course, another explanation for the lack of debate is that few are unhappy with macroeconomic performance under the current regime.

Mandates in Inflation-Targeting Regimes

New Zealand in 1990 became the first country to establish a formal inflation-targeting regime. Canada followed in 1991, the United Kingdom in 1992, and Australia and Sweden in 1993. Subsequently, Finland and Spain adopted inflation targeting (before becoming members of the European Monetary Union), and in the last few years several developing countries have adopted this approach. Although the European Central Bank (ECB) does not identify itself as an inflation-targeting regime, the Maastricht Treaty set price stability as the ECB's primary objective and the ECB has set an explicit numerical target for inflation.

What Is an Inflation-Targeting Regime? Inflation-targeting regimes generally identify price stability as the primary objective, usually in the context of a hierarchical mandate. They set an explicit numerical target for inflation and set a period over which any deviation of inflation from its target is to be eliminated, although some regimes provide escape clauses and other flexibility related to the pace of return to price stability.

The inflation target is sometimes set as a point and sometimes as a range. In most cases, the inflation objective is set for a measure of overall consumer price inflation, the point or midpoint of the ranges is generally around 2 percent, and the ranges (where employed) are generally 2 percentage points wide—typically 1 percent to 3 percent. The time period prescribed for return to the inflation target following departures is sometimes explicit and sometimes not, generally in the range of 18 months to 2 years.

Examples of Inflation-Targeting Regimes. In New Zealand, the first inflation-targeting regime, the numerical target is set jointly by the Minister of Finance and the Governor of the central bank and is currently a range of 0 percent to 3 percent, the widest of any of the ranges in inflation-targeting regimes. New Zealand is quite well-known for establishing performance contracts for government officials, and this approach is followed in the law governing the operation of the central bank: The statute allows the Governor to be dismissed if inflation performance is inadequate.

The Bank of Canada operates under the vaguest legal mandate among inflation-targeting central banks. Its statute requires it to regulate "credit and currency in the best interests of the economic life of the nation." Despite the absence of a precise legal mandate, the details of the Bank's monetary policy objectives are reached by agreement between the Bank and the Department of Finance. This agreement has set price stability as the principal objective

for monetary policy. To implement this objective, the agreement sets the range for inflation as 1 percent to 3 percent and identifies the midpoint as the explicit target.

The Reserve Bank of Australia has a mandate most closely resembling ours, though it is even broader and more open-ended. Their legislative mandate is "to [promote] stability of the currency of Australia;…[maintain] full employment in Australia; and…[foster] economic prosperity and welfare of the people of Australia." The explicit inflation target, 2 percent to 3 percent, is set by the central bank and applies to the average inflation rate over a business cycle. Although Australia is counted among inflation-targeting countries, it has a dual mandate rather than a hierarchical one. Indeed, it is a model for the combination I prefer: an explicit inflation target within a dual mandate.

The mandate in the United Kingdom is hierarchical. Article 11 of the Bank of England Act sets the objectives for monetary policy as follows: "to maintain price stability" and "subject to that, to support the economic policy of Her Majesty's Government, including its objectives for growth and employment." The explicit target, set by the Chancellor of the Exchequer (the equivalent of the Minister of Finance in many countries or the Secretary of the Treasury in the United States), is currently 2.5 percent and the target is for retail prices excluding mortgage interest payments. The Governor of the Bank of England must write a letter to the Chancellor if inflation deviates by more than 1 percentage point from the target.

The ECB does not view itself as an inflation-targeting central bank. However, the Maastricht Treaty—the equivalent of the statute establishing the objectives for a central bank—identifies price stability as the principal objective in the context of a hierarchical mandate. Article 105 of the Maastricht Treaty states that "the primary objective of the [European System of Central Banks (ESCB)] shall be to maintain price stability. Without prejudice to the objective of price stability, the ESCB shall support the general economic policies in the community with a view to contributing to the objectives of the Community laid down in Article 2." The objectives mentioned in Article 2 include "sustainable and non-inflationary growth," a "high level of employment," and "raising the standard of living" among member states. The ECB's Governing Council sets the explicit numerical inflation target. This is currently set with an explicit ceiling of 2 percent and an implicit lower bound of 0 percent. This is the case of a range rather than a point, with no preference stated for the midpoint.

The Evolution of Inflation-Targeting Regimes. Over their short history, inflation-targeting regimes have evolved to give central banks greater flexibility in conducting monetary policy. Mervyn King calls regimes which take no account of output gaps (where the coefficient on the output gap is zero in the loss function) "inflation nutters."[4] That language suggests that entirely ignoring output stabilization is now viewed as an extreme position and not as a desirable option for central banks. Lars Svensson argues that there has, in fact, been a convergence toward "flexible inflation targeting"—meaning inflation-targeting regimes that in practice take into account deviations in both output and inflation from their respective targets.[5] Such an evolution has brought many inflation-targeting regimes closer in practice to a dual mandate regime.

THE CASE FOR A DUAL MANDATE

The appropriate goals for monetary policy depend on the structure of the economy and the preferences of the citizenry. My support for a dual mandate reflects my views about the structure of the economy and about the public's preferences. These can be summarized as follows:

1. Low and stable inflation ("price stability") is essential to good macroeconomic performance and hence should be an objective of macroeconomic policy.
2. The central bank is uniquely responsible for the inflation rate in the long run.
3. Monetary policy can make some contribution to lowering the variability of output relative to potential.
4. The public desires both low and stable inflation and a low variability of output relative to potential.

The first two points, of course, are shared by most central bank mandates. The case for the dual mandate includes the third and fourth points.

Price Stability as a Policy Objective

As I noted at the outset, it is widely agreed that low and stable inflation is desirable. Several costs

[4] King (1997).

[5] Svensson (1999).

of high and variable inflation have been identified. These costs typically arise from distortions in economic decisionmaking arising from high or variable inflation rates and result in lower levels of output than would otherwise be the case. I won't elaborate in detail about these costs here, because I take as a starting point the agreement that price stability is an important, if not the singular, objective for monetary policy.[6] But the key point is that price stability is not an end in itself; it is important because it contributes to a higher level of output and perhaps faster growth in output.

Monetary Policy and Inflation

Few economists would disagree that inflation is, as Milton Friedman taught us long ago, always and everywhere a monetary phenomenon. This was earlier interpreted as a statement about a tight relationship between money growth (controlled by the central bank) and inflation. Today, it is recognized that even if the relationship between money growth and inflation has weakened, perhaps because of financial innovations, central banks can achieve their inflation targets by adjusting their preferred instrument, typically some short-term interest rate. Hence, monetary policy still determines the rate of inflation in the long run. While it is also well understood that supply shocks—such as abrupt changes in the price of energy or food unrelated to the overall balance between aggregate demand and supply—can result in short-run changes in inflation, such changes in inflation can persist only if central banks accommodate them. Central banks therefore must accept full responsibility for inflation in the long run and have the tools to achieve price stability.

The Ability to Stabilize Output

While monetary policy can achieve a long-run inflation target, economic theory suggests that it cannot affect the level of output or its growth rate in the long run, other than by maintaining low and stable inflation. Therefore, the objective of price stability should be assigned to monetary policymakers, but the objective of high and rising living standards should not be. On the other hand, the Congress and the Administration have many opportunities to affect the level and growth in potential output—including the size of the structural budget deficit relative to output, the details of the tax code, and the composition of government spending.

Whether central banks should accept responsibility for stabilizing output relative to potential is more controversial. Milton Friedman, for example, has always questioned the ability of central banks to stabilize output relative to potential and worried that attempting to do so could be counterproductive, given the "long and variable lags" between policy actions and the economic effects. Economists agree that monetary policy cannot "fine tune" the economy to ensure that the full employment objective is continuously maintained. However, a considerable amount of research supports the contention that monetary policy can reduce the variability of output around its full-employment level.[7]

Preferences

Households and businesses are presumed to prefer low and stable inflation to high and variable inflation. But they also prefer high and rising real income per capita and output that is consistently close to the economy's maximum sustainable level of output. This is often expressed in terms of a loss function where the loss to society is expressed as a weighted average of squared deviations of inflation from its target and of output from its potential level. The squaring of the deviations ensures that deviations on either side of the target are treated equivalently as losses.[8] The weights, a and $1-a$, indicate the relative intensity of the public's distaste for deviation from their preferred rates of inflation and output. The loss to society, L, can be expressed as

$$(1) \qquad L = a(\pi - \pi^*)^2 + (1-a)(y - y^*)^2,$$

where π is the rate of inflation, π^* is the target rate of inflation, y is the level of output, and y^* is the target level of output or potential output.

---

[6] For a discussion of the case for price stability, see Fischer (1996).

[7] Much of this research involves simulations of empirical macroeconomic models with alternative policy rules. The simple policy rules used in these exercises mimic the systematic aspects of the response of discretionary policy to changes in the macroeconomy. For example, Taylor has shown that Federal Reserve policy actions in recent years have been broadly similar to what a simple policy rule would have prescribed. These exercises therefore demonstrate the ability of simple rules—and by extension, discretionary monetary policy more broadly—to both reduce the variability of output and achieve a long-run inflation target. See Taylor (1999) for a series of papers involving simulations of models with various policy rules.

[8] Squaring the deviations from target values also penalizes larger deviations more than proportionately compared with smaller ones.

Tradeoff Between Inflation and Output Variability

Although it is possible in principle to achieve price stability and full employment simultaneously, an inevitable tradeoff between the variability of output and the variability of inflation exists.[9] This tradeoff is most obvious in the case of a supply shock, for example an abrupt increase in the price of oil. An adverse supply shock typically raises inflation and lowers aggregate demand (by reducing the purchasing power of consumers), thereby moving inflation up and output down. This gives rise to a well-known dilemma for monetary policy: Should monetary policy ease to reduce the decline in output or tighten to counter the rise in inflation? The structure of the economy is such that the quicker monetary policy tries to return inflation to its target (to reduce the variability of inflation), the greater the variability in output.

The choice of a hierarchical versus a dual mandate may be the most important consideration determining where a country ends up along this tradeoff. That is, countries with hierarchical mandates are more likely to end up with lower inflation variability at the expense of higher output variability.[10] A dual mandate, therefore, provides flexibility for the central bank to select the point along this tradeoff that matches the public's preference.

Focusing Policymakers on What They Can Achieve

An argument often made for a hierarchical mandate is that identifying price stability as the sole or primary objective focuses monetary policy on what it can achieve and, thereby, reduces political pressure to pursue goals that are not achievable. To be sure, wide agreement exists that central banks can achieve low and stable inflation and should be held accountable for doing so. There is also agreement that central banks cannot raise the level or growth of potential output (other than through achieving their price stability objective) and should therefore not be held accountable for these dimensions of macroeconomic performance. But the public also cares about the variability of output relative to potential, and the conduct of monetary policy inevitably will affect output variability. Therefore, in my view, the importance of keeping monetary policy focused on what it can do supports the case for a dual mandate that explicitly recognizes both price stability and output stabilization as objectives for monetary policy.

Is There a Single Long-Run Objective?

It is sometimes argued, however, that price stability can be the only objective for monetary policy in the long run, again placing price stability on a higher plane than full employment. In the long run, theory holds that the economy gravitates to full employment by self-equilibrating forces, principally through the effects of price flexibility. That is, if the economy is operating at a level of output below full employment, the price level will tend to fall, and at least for a given value of the nominal money supply, this will tend to stimulate aggregate demand. Over time, this process will raise aggregate demand to a level consistent with full employment. Hence, policymakers do not have to be concerned with full employment in the long run, leaving price stability as their unique long-run concern.

But this view is misleading in a couple of respects. First, monetary policymakers should be concerned about *two* long-run properties of the economy. One is price stability and the other is the variability of output around full employment. Policy has to be judged by its success in both dimensions. Second, policy is made in the short run, not the long run. The speed of return of output to its potential level is influenced by policy decisions and cannot be treated with indifference. It may just take too long and waste too many resources in the interim to rely on the self-equilibrating forces of the economy. Policymakers will therefore have to take into account, in practice, both objectives in their policy actions.

Still, a meaningful difference in the interpretation of the two objectives remains. A central bank can achieve the inflation target, with considerable precision, in the long run, meaning on average over a period of years. But it cannot be expected to maintain the economy continuously at full employment. The full employment objective might therefore be better interpreted as an output stabilization objective. It instructs the central bank to work to reduce the variability of output around its full employment level.

The Need for Flexibility

The purpose of a hierarchical mandate is to impose constraints on the operation of monetary policy, constraints that proponents believe enhance

[9] Levin, Volker, and Williams (1999) analyze the output-inflation volatility frontier in four different macroeconomic models.

[10] For some supporting evidence, see Cecchetti and Ehrmann (1999).

credibility, focus policy on what monetary policy can achieve, and reduce political pressures for policy to aim at impossible-to-achieve and potentially destabilizing output goals, such as a level of output above the economy's maximum sustainable rate. However, those same constraints might interfere with the pursuit of other legitimate objectives of monetary policy, specifically with policy adjustments to reduce the variability of output around potential output.

Most inflation-targeting regimes explicitly recognize that returning inflation to its target too rapidly following some departure could result in excessive variability of output. The solution has been to encourage a gradual return to the inflation target by explicitly or implicitly setting a policy horizon over which policymakers commit to return inflation to its target. Setting fixed horizons for the return to the inflation target, independent of the size or the nature of the shock, clearly reduces the flexibility of monetary policy. To be fair, many regimes explicitly note that the policy horizons need not be fixed or include escape clauses that would allow greater flexibility, for example, in response to a supply shock. But setting a policy horizon is intended to, and does, constrain policy responses. It may therefore interfere with an appropriate balancing of the full-employment and price-stability goals. This is especially the case if the mandate is hierarchical, where other objectives can be pursued only if the inflation objective is achieved.

The Taylor Rule and the Dual Mandate

The Taylor rule is a useful characterization of U.S. monetary policy.[11] According to the Taylor rule, monetary policymakers should adjust the target for the short-term interest rate in response to deviations of output and inflation from their respective targets and in response to changes in inflation. It is therefore well aligned with a dual mandate. The Taylor rule can be written as

$$(2) \qquad R = r^* + \pi + c[(y/y^*) - 1] + d(\pi - \pi^*),$$

where R is the target nominal policy rate, r^* is the equilibrium real level of the policy rate (consistent with price stability and full employment), y is output, y^* is the level of potential output, π is inflation, π^* is the target for inflation, and c and d are the parameters that describe the response of the policy rate to deviations of output and inflation from their respective targets.

The Taylor rule is consistent with the loss function described in equation (1) because the rule prescribes an adjustment of the federal funds rate in response to the deviations from target values that are presumed to give rise to costs to society. The Taylor rule also helps to make the point that policymakers can operate with an output stabilization goal and still ensure that inflation is, on average, consistent with the inflation target in the long run.[12]

A strict version of inflation targeting would be one based on a loss function, taking into account only losses associated with inflation deviating from its target. In terms of equation (1), that would mean that $a = 0$—that is, no costs were associated with deviation of output from its potential level. Based on a simple model, we could derive an optimal policy response to shocks, based on this loss function and the more general one where costs are assigned to deviations from both price stability and full employment. If policy is set assuming $a = 0$ (as would be the case in a very strict inflation-targeting regime), it is clear that policy will be suboptimal if the true loss function does assign a cost to deviating from full employment.[13]

The more difficult question is: How suboptimal are inflation-targeting regimes that recognize the costs of excessive output variability, but nevertheless constrain monetary policymakers from responding to deviations of output from its target, except when the inflation target has already been met or when policymakers can project that it will be met in a reasonable period? In my view, such regimes are likely to remain suboptimal, compared with a more flexible dual-mandate regime.

Transparency, Accountability, and the Dual Mandate

As I noted, transparency about monetary policy requires a full and accurate account of the objectives. But pretending that inflation is the only objective, while taking account of output variability in practice, only makes for less-transparent policy and ensures that the central bank will have difficulty communicating the rationale for its policy actions.

---

[11] Taylor (1993).

[12] A Taylor rule could also be consistent with a regime that set inflation as the sole objective. In this case, the output gap terms would be viewed as a predictor of inflation, allowing policymakers to take preemptive action to contain the threat of higher inflation, in addition to directly responding to higher inflation itself.

[13] See Kim and Henderson (2000).

I remember the first conference I attended after joining the Board of Governors. Two foreign central bankers—each from inflation-targeting countries—lectured me about how "good" central bankers acted in public. They each told me that a disciplined central banker would never admit to having a stabilization objective and never admit that there was a cost of lowering inflation. Such admissions, they warned, would only undermine the public's confidence in a central banker's commitment to price stability. I responded that this lesson in central banking surprised me. I would not have thought obfuscating about policy objectives or the way monetary policy affects the economy would have enhanced the credibility of a policymaker. I still don't.

THE CASE FOR AN EXPLICIT INFLATION TARGET

An explicit inflation target would, in my view, give added precision to an already mandated objective. Three of the arguments for inflation-targeting regimes, it seems to me, also support an explicit inflation target within a dual mandate. First, an explicit inflation target would improve the transparency and accountability of monetary policy. Second, it might help, at the margin, to anchor inflation expectations. Third, it would help to institutionalize recent good monetary policy. It would also, in my view, make the decisionmaking process more coherent. Indeed, moving in this direction would extract most of the benefits of an inflation-targeting regime without suffering the loss of flexibility inherent in its hierarchical mandate.

Improving Transparency and Accountability

Transparency is an important goal for monetary policymakers for two reasons. First, a more transparent policy may be a more effective one. Monetary policy works both through the setting of a target short-term nominal interest rate and by the expectations policymakers induce in the markets, wittingly or unwittingly, about the course of future policy. To the extent that market participants correctly anticipate future policy moves, long-term interest rates will move in response to expectations of future moves in short-term rates, in effect, speeding the response of aggregate demand to monetary policy. Second, central bank independence has to be balanced by accountability. More precise goals increase both the transparency and the accountability of monetary policy.

Anchoring Inflation Expectations and Increasing Credibility

Anchoring inflation expectations at the targeted rate of inflation is an important goal for monetary policy. However, whether an explicit inflation target or a formal inflation-targeting regime would boost the credibility of the central bank's commitment to price stability is more debatable. In my judgment, credibility is primarily earned by performance. In addition, there is little empirical evidence to suggest that either explicit inflation targets or inflation-targeting regimes lower the cost of disinflation by directly lowering inflation expectations. Still, at the margin, such a target might enhance the ability of the Fed to anchor inflation expectations and perhaps also enhance the ability of the Fed to pursue its stabilization objective without undermining the public's confidence in its commitment to low inflation.

Institutionalizing Good Monetary Policy

Most observers would rate monetary policy in the United States over the last two decades as very good. Good policy, in turn, depends on the combination of a well-defined mandate, a disciplined strategy for achieving the mandate, and the quality of the FOMC and its staff working to implement the strategy. The Chairman has considerable influence over the policy outcome, disproportionate to the one vote he or she casts, so that the quality of the chairman is especially important. The Federal Reserve has been fortunate to have strong leadership for many years under both Paul Volcker and Alan Greenspan. A more fully articulated mandate could help ensure that policy remains well focused and disciplined as the leadership of the Fed changes.

Facilitating the Policymaking Process

The Fed staff routinely shows policymakers the prescriptions from several Taylor-rule-type policy reaction functions, and such a rule is explicitly incorporated into the Fed's large-scale model used for policy analysis. The staff has never asked FOMC members about their preferences for a numerical inflation target and instead often employs the target that John Taylor used in the policy rule he introduced in 1993. I have great regard for John, but this seems to be a rather unusual way for a policy rule to be run at the Fed.

More important, if the target is implicit, and

therefore imprecise, members of the FOMC inevitably will each make policy decisions in pursuit of different inflation targets. It seems to me that the internal discussion of policy would be more coherent if policymakers agreed in advance on an inflation objective. There would still be differences in policy preferences at FOMC meetings due, for example, to different views about the economic outlook, different views about the structure of the economy, and different views about the sustainable level of the unemployment rate or the maximum sustainable rate of growth in output. But, provided that all FOMC members agreed to seek the objective chosen by the majority, an explicit target would prevent members from pulling in different directions because of different inflation objectives.

Retaining Flexibility with the Dual Mandate

The key issue for me is whether setting an explicit inflation target would reduce the flexibility of policymakers to pursue a dual mandate and select the preferred point along the tradeoff between output and inflation variability. That is, would making the p^* in the Taylor rule explicit inevitably also lead to a change in the relative responses to deviations from the output and inflation targets (the c and d parameters in equation (2))? Specifically, would implementing an explicit inflation target inevitably also raise the response parameter on the inflation gap relative to that on the output gap? In my view, the answer is that this need not be the case, but I agree that there is some risk of this outcome. It seems to me, however, that it is less likely if the move to an explicit inflation target is taken in the context of a reaffirmation of the dual mandate.

If It Ain't Broke, Don't Fix It

As I noted above, most observers believe that monetary policy has been excellent in recent years. Why, then, change the underlying framework for policy, especially when this change is not likely to have much of an influence on the conduct of monetary policy?

The case for moving to an explicit inflation target within our current dual mandate is that it extracts most of the potential benefits of inflation-targeting regimes without necessarily reducing the flexibility of policy. I believe that the Fed remains more responsive to deviations of output from its target than most inflation-targeting central banks, even allowing for the convergence that has been under way. I believe we could retain that flexibility with an explicit inflation target because of our historical commitment to a dual mandate and because there would be no presumption that the fundamental strategy for conducting monetary policy, summarized by the Taylor rule, would change. At the same time, I believe transparency and accountability are appropriate goals for monetary policymakers and that an explicit inflation target would contribute to each of these goals, even if it would not have had much effect on the actual course of policy in recent years.

Should the Full Employment Objective Be Explicit Also?

A natural question in the case of a dual mandate is whether both objectives should be made explicit. Some members of the Congress, for example, might encourage the Fed to adopt an explicit objective for full employment to balance an explicit inflation objective. Although I support an explicit objective for inflation, it would not be constructive, in my view, to set an explicit numerical target for full employment.

The central bank is capable of achieving an inflation objective, at least on average over a period of years. In contrast, if we define full employment in terms of a threshold for the unemployment rate consistent with maximum sustainable employment, the central bank has no choice about what this threshold should be.[14] It is determined by the structure of the economy, including the effectiveness of institutions and markets in matching vacancies and unemployed workers, and by policies, such as the levels of unemployment compensation and minimum wage rates.

Because institutions and markets evolve and labor market policies change, it would be inappropriate to set a fixed numerical objective for full employment. That does not rule out central banks being more transparent about their estimate of the unemployment rate that is consistent with maximum sustainable employment, though no central bank has ventured into this territory. That is perhaps not surprising. Indeed, it would be difficult to do so because there is considerable debate—both inside and outside central banks—about the usefulness of a "full employment rate of unemployment" as a

[14] This threshold is often referred to as the non-accelerating inflation rate of unemployment or the NAIRU.

guide to policy and, even among those who believe this is an important and useful concept, about what that rate may be today and how it tends to change over time. So the issue here is not simply one of transparency and accountability.

In addition, political considerations may be at work. At times, an increase in the unemployment rate may be required to sustain stable, low inflation. Rare is the leader of a monetary policy committee who relishes going to the legislature and reporting that the central bank is concerned that too many people have jobs.

The uncertainty about the threshold unemployment rate also suggests a differing degree of intensity in the response of monetary policy to deviations of inflation and output to their respective targets. That uncertainty derives from our inability to precisely pin down two key parameters—the threshold unemployment rate and the trend rate of growth in potential output. As a result, there is a subtle difference between the two objectives in the dual mandate. One objective—price stability—can be well defined and is fully under the control of monetary policymakers, at least over a period of time. The other—full employment—is set by the structure of the economy, not policymakers, and cannot be precisely pinned down numerically at any point in time and can vary over time.

PRACTICAL CONSIDERATIONS

Let us now assume that a decision has been made to move to an explicit inflation target. Several steps would be required to implement such a decision. First, we would have to decide who would set the explicit target. Second, a specific price index would have to be selected to serve as the basis for the inflation target. Third, we would have to decide whether the target should be a point or a range. Fourth, we would have to decide on the level of the point or the values that establish the range.

Who Should Set the Inflation Target?

Among inflation-targeting regimes, there is a mix of practices with respect to who sets the numerical target for inflation. In almost all cases, the government identifies price stability as a target, either as the single target or as part of a hierarchical mandate. In about half the cases, the explicit numerical target for inflation is set by the government, typically the finance ministry, generally in consultation with the central bank; in about half the cases, the target is set by the central bank, often in consultation with the finance ministry.

The United States already has a mandate that includes price stability. The existing law therefore seems to be compatible with the Federal Reserve setting an explicit numerical inflation rate consistent with the notion of price stability. Thus, the FOMC could move in this direction without any amendment to the Federal Reserve Act. Perhaps the setting of an explicit inflation target by the Federal Reserve might be analogous to its earlier setting of numerical money growth ranges. In 1978, the Congress instructed the Fed to report an objective for money growth. The FOMC then decided unilaterally on the numerical ranges. In 2000, the Congress amended the Federal Reserve Act to remove the requirement that the Fed report to the Congress on the ranges for monetary aggregates. In the case of an inflation objective, the Congress has already imposed a price-stability objective. All the FOMC would be doing is communicating back to the Congress and the public its interpretation of that objective.

Nevertheless, such a move would likely be interpreted as an important change in the conduct of monetary policy. Consequently, if the FOMC desired to move in this direction, extensive prior consultations would be appropriate, especially with the Congress, but also with the Administration. It might also be constructive to get feedback from a wider audience on a proposal to set an explicit inflation target before proceeding in this direction.

It would, however, be problematic for the United States to follow the approach in many inflation-targeting countries of having the ministry of finance (in our case, the Treasury Department) set the inflation target. Such an approach is more comfortable politically in a parliamentary system, in which the legislative and executive branches are always controlled by the same party or coalition. It seems unlikely that the Congress would delegate this responsibility to the Treasury Department. Even regular consultation with the Treasury Department about a specific goal under the approach I have mentioned here might raise questions unless the Congress explicitly authorized it.

Choosing a Specific Price Index

Most broad measures of inflation move roughly in concert over time. Nevertheless, the differentials among the most widely used aggregate measures of inflation generally are not stable over long periods. Therefore, it does make some difference which

measure is selected for the target. All inflation-targeting central banks use a measure of consumer price inflation for their target. In the absence of a complete model of the cost of inflation to guide the choice, the use of consumer price measures seems appropriate because these measures are the most relevant to the calculation of real income for households, because the welfare of households is ultimately the goal for monetary policy, and because measures of consumer price inflation are often the most visible of inflation measures.

In the United States, this would mean a choice between the consumer price index (CPI) and the price index for personal consumption expenditures. The CPI is the most widely recognized measure of consumer price inflation, but the Federal Reserve has expressed the view that the measure based on personal consumption expenditures has advantages as a measure of trends in consumer price inflation. Setting the inflation rate in terms of the price index for personal consumption expenditures would further elevate this measure in the public consciousness.

A related decision is whether to set the target in terms of overall inflation or in terms of core inflation—that is, consumer price inflation net of the direct influence of movements in energy and food prices, which tend to be volatile. Focusing on core inflation might increase the likelihood that monetary policy would respond to underlying inflation developments that promised to be persistent, and not on transitory changes in inflation. Nevertheless, most inflation-targeting regimes set their explicit inflation target in terms of overall inflation. As long as policy is focused on achieving a longer-run inflation outcome, as opposed to meeting the target each quarter or even each year, a target for overall inflation can work well because swings in energy and food prices that lead to shorter-run inflation blips tend to dissipate or wash out over longer periods. In this case, it would also be useful, as is the practice at some inflation-targeting central banks, to monitor movements in core inflation as a guide to whether current monetary policy is well aligned with the longer-term objective for overall inflation. Weighing the pros and cons, policy might be better and the communication to the public might be most clear if the target itself was expressed in terms of the core inflation rate.

Point or Range

Practice differs among inflation-targeting central banks, with about half setting their targets as points and the other half as ranges. Among those who adopt ranges, some also identify the midpoint of the range as the specific target. I prefer a point target. This provides a more precise anchor for inflation expectations and a more specific target for monetary policymakers. If a range is selected, questions will arise about the differing implications of movements of inflation inside the range and outside the range and, in the absence of explicitly identifying the mid-point as the target, about where within the range policymakers would prefer inflation to gravitate. If there is indifference about movements inside the range, this will tilt policy toward allowing greater variability of inflation.

What Level?

If the objective is price stability, this may seem like a pointless question. But it is an important issue for two reasons. First, assuming that the official indexes are biased upward, true price stability will be achieved at some positive rate for measured inflation. Second, there are considerations that favor allowing some small positive rate of true inflation, rather than true price stability.

In 1995, the Boskin Commission estimated that the measurement bias for the CPI was between 0.8 and 1.6 percentage points. Since then, the Bureau of Labor Statistics has made a series of methodological improvements in the CPI, lowering the bias. In 1999, the Congressional Budget Office estimated that the cumulative effects of these improvements had been to reduce the measurement error to a point estimate of 0.8 percent. Therefore, in terms of the CPI, if this were the only consideration, the target could be set at 1 percent to allow for measurement error (and also a round number). This also suggests that the target should not be fixed in time, but should be adjusted over time to take into account changes in the estimated bias.

There are a couple of reasons to consider setting an inflation target above the level consistent with true price stability. If there is resistance to declines in nominal wages, a small amount of wage inflation may enhance the flexibility of real wages and facilitate the movement to equilibrium in the labor market.[15] Whether this requires positive price inflation will depend on the variance of the wage changes, the degree of nominal wage rigidity, and the rate of growth of labor productivity. In addition, a positive rate of inflation increases the flexibility of mone-

[15] See Akerlof, Dickens, and Perry (1966).

tary policy by allowing policymakers to drive real interest rates below zero. Particularly in light of the latter consideration, I would set the inflation target at what I refer to as true price stability plus a small cushion. Specifically, for the CPI, I would set it at 2 percent, 1 percentage point for the measurement error and 1 percentage point for the cushion.[16] This would correspond roughly to a 1.5 percent inflation rate for the price index for personal consumption expenditures, based on recent differentials of this measure relative to the inflation rate for the CPI. The precise magnitude of the cushion should also be subject to adjustment over time to reflect ongoing research about its optimal size.

Time Horizon

There should be no fixed time horizon for returning inflation to its target when deviations occur. Such a horizon would be arbitrary, in some cases might not be credible, and potentially would constrain the ability of the FOMC to pursue the dual mandate. On the other hand, it would be useful to use a measure for inflation that smoothes over transient shocks and that is not subject to potential problems with seasonal adjustment. For that reason, I would favor a year-over-year measure of the inflation rate—for example, a 12-month CPI or price index for personal consumption expenditures—for evaluating performance relative to the target.

CONCLUSION

It is widely agreed that price stability is an important objective of monetary policy and that central banks should be held responsible for that objective. That said, central banks differ over whether they should set an explicit inflation target and whether they should acknowledge and take responsibility for other objectives, specifically full employment or output stabilization.

An explicit inflation target would give added precision to an already mandated objective and would thereby improve the transparency of and accountability for monetary policy. Moving to an explicit inflation target would, in my view, be consistent with the current statute governing the objectives of monetary policy in the United States and would, in turn, require no fundamental change in the current Federal Reserve strategy for implementing monetary policy.

The most important question that has to be addressed in order to assess the costs and benefits of a move in this direction is whether it could be accomplished without reducing the flexibility the Fed now has to pursue a dual mandate. In my view, if the explicit mandate is set in the context of a reaffirmation of the dual mandate, the flexibility now enjoyed by U.S. monetary policy will likely be maintained. A second question is whether moving in this direction would matter much for the conduct of monetary policy in the United States. I believe the answer to that question is that it would not matter much today, with the current Chairman and the current FOMC. But moving in this direction would provide some greater assurance of continuity in policy. In addition, it would have the virtue of enhancing transparency and accountability.

REFERENCES

Akerlof, George; Dickens, William and Perry, George. "The Macroeconomics of Low Inflation." Brookings Papers on Economic Activity, 1996, *1*, pp. 1-59.

Bernanke, Ben S.; Laubach, Thomas; Mishkin, Frederic S. and Posen, Adam S. *Inflation Targeting: Lessons from the International Experience.* Princeton, NJ: Princeton University Press, 1999.

Cecchetti, Stephen G. and Ehrmann, M. "Does Inflation Targeting Increase Output Volatility? An International Comparison of Policymakers' Preferences and Outcomes," in Klaus Schmidt-Hebbel, ed., *Monetary Policy: Rules and Transmission Mechanisms,* Proceedings of the Third Annual Conference of the Central Bank of Chile, 1999. <www.bcentral.cl>.

Fischer, Stanley. "Why Are Central Banks Pursuing Long-Run Price Stability," in *Achieving Price Stability*: A symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, WY, August 1996, pp. 7-34.

Fuhrer, Jeffrey C. and Sniderman, Mark S., eds., "Monetary Policy in a Low Inflation Environment." Summary of a conference sponsored by the Federal Reserve Banks of Boston, New York, Cleveland, Richmond, Atlanta, St. Louis,

[16] See Fuhrer and Sniderman (2000) for a series of papers on the implications of a zero nominal bound for the nominal interest rate for setting inflation targets. The paper by Reifschneider and Williams in that volume provides some support for the 1 percentage point cushion I have suggested. They conclude: "[I]n very low inflation environments… the zero bound could prove to be a significant constraint on policy… With the effectiveness of open market operations diminished at times, the economy would likely experience a noticeable increase in the variability of output and employment, particularly if policymakers were to pursue an inflation target of 1 percent or below."

and Minneapolis, and the Board of Governors of the Federal Reserve System, 18-20 October 1999. *Journal of Money, Credit, and Banking*, November 2000, *32*(4, Part 2), pp. 845-69.

Gramlich, Edward M. "Inflation Targeting." Speech delivered before the Charlotte Economics Club, Charlotte, NC, 13 January 2000. < www.federalreserve.gov/boarddocs/speeches/2000/20000113.htm > .

Greenspan, Alan. Statement before the Subcommittee on Economic Growth and Credit Formulation of the Committee on Banking, Finance, and Urban Affairs, U.S. House of Representatives, 22 February 1994.

Kim, Jinill and Henderson, Dale. "Exactly How Suboptimal Are Inflation Targeting and Nominal Income Targeting." December 2000. < www.people.virginia.edu/ ~ jk9n/hk3001219.pdf > .

King, Mervyn. "Changes in U.K. Monetary Policy: Rules and Discretion in Practice." *Journal of Monetary Economics*, June 1997, *39*, pp. 81-97.

Levin, Andrew T.; Volker, Weiland and Williams, John C. "Robustness of Simple Monetary Policy Rules Under Model Uncertainty," in John B. Taylor, ed., *Monetary Policy Rules*. Chicago: University of Chicago Press, 1999.

Svensson, Lars E.O. "How Should Monetary Policy Be Conducted in an Era of Price Stability?" in *New Challenges for Monetary Policy*: A symposium sponsored by the Federal Reserve Bank of Kansas City, August 1999, pp. 195-259.

Taylor, John B. "Discretion Versus Policy Rules in Practice." Carnegie-Rochester Conferences Series on Public Policy, December 1993, *39*, pp. 195-214.

_____, ed. *Monetary Policy Rules*. A National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press, 1999.

Volcker, Paul. "Can We Survive Prosperity?" Speech given at the Joint Meeting of the American Economic and American Finance Associations, San Francisco, CA, 28 December 1983.

Equity Financing of the Entrepreneurial Firm

Frank A. Schmid

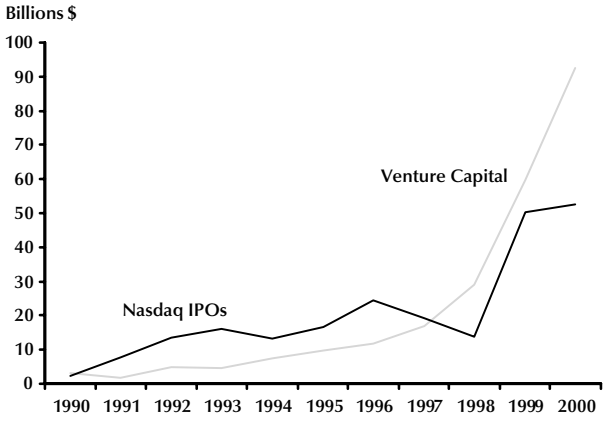
An entrepreneur is an individual with a project blueprint and limited wealth. If launching the project requires expenses that exceed the entrepreneur's initial wealth, he needs outside financing. Entrepreneurs differ from "hired management" in that they are indispensable for the firm's day-to-day operations. This is because entrepreneurs add value to companies perpetually, rather than by handing over the project blueprints.

Outside financing is fraught with the problem of asymmetric information between the entrepreneur (who is a firm insider) and the (outside) investor. Asymmetric information between management and investor is considered the most significant problem in corporate finance.[1] Typically, the problem of asymmetric information is modeled in finance literature as one that pertains to the use of free cash flow by management or to management's project choice.[2] Asymmetric information about the use of free cash flow can take on a variety of forms. First, cash flow might be unobservable. In this case, the diversion of free cash flow for personal use by management goes unnoticed by the investor. If cash flows are unobservable, the outside equity holder has no bargaining power over the allocation of free cash flow to dividend payments. Second, management's use of free cash flow might be observable, but not verifiable. This is when the outsider can observe management directing free cash flow to its own benefit, but cannot verify these actions in court. Third, management's actions might be observable and verifiable, but compliance might not be enforceable. Examples of non-enforceability are cases where it is prohibitively costly for investors to go to court, or where court rulings are rendered worthless because the culprit is subject to limited liability or has limited wealth.[3]

In spite of the problems of asymmetric information outlined above, outside equity financing of the entrepreneurial firm has achieved a rapid increase over the past decade (see Figure 1). Venture capital funds, which finance privately held start-ups, raised a record $92.3 billion in 2000. This is a thirty-fold increase relative to 1990. At Nasdaq, initial public offerings raised an all-time high of $53.6 billion in 2000, which is 24 times as much as in 1990.

Figure 1

Venture Capital Raised and Nasdaq IPOs



NOTE: Annual observations; last observation: 2000.
SOURCE: Venture Economics & National Venture Capital Association, <http://www.ventureeconomics.com>; Nasdaq, <http://www.marketdata.nasdaq.com>.

This article analyzes equity financing of the entrepreneurial firm against the background of observable but non-verifiable cash flow. The study covers three organizational forms: the limited partnership, the private corporation, and the public corporation. The legal type of the firm determines the outside equity financing options that are available to the entrepreneur. By definition, initial public offerings are available to the (henceforth) public corporation only, whereas venture capital is possible for both the limited partnership and the privately held corporation.

The analysis shows that venture capital financing is superior to public offerings when the entrepreneur has low initial wealth relative to the size of the project and is equivalent otherwise. This result highlights the importance of private equity in financing entrepreneurial enterprises. The Gramm-Leach-Bliley Act of 1999 allows banks to expand

Frank A. Schmid is a senior economist at the Federal Reserve Bank of St. Louis. William Bock and Judith Hazen provided research assistance.

[1] See Hart (1995), Jensen (2000), and Myers (2000).

[2] See Zingales (2000).

[3] For causes of lack of observability, verifiability, or enforceability and their consequences for contractual arrangements, see Hart (1995) or Milgrom and Roberts (1992).

the scope of their activities in this arena. The law allows financial holding companies to provide equity financing to nonfinancial firms for up to ten years. In particular, the act defines a framework in which financial holding companies can sponsor private equity funds that provide venture capital to entrepreneurial firms. While it is not the purpose of this article to study the consequences of the Gramm-Leach-Bliley Act on venture capital financing, the analysis suggests that venture capital is a significant financing instrument. The Gramm-Leach-Bliley Act helps improve the supply side in the venture capital market by broadening the spectrum of institutions that are allowed to provide private equity to firms that do not (or not at this stage) seek to raise capital in an initial public offering.

The following analysis uses a simple model of the entrepreneurial firm as proposed by Myers (2000). While the analysis of the limited partnership draws heavily on Myers, the examination of the corporation differs from Myers most significantly in its simplicity and its focus on the most basic differences among the organizational forms. The model shows that private corporations face less stringent financing constraints than limited partnerships or public corporations. The result emphasizes the importance of the private corporation as a legal form. While the model helps explain the choice of the legal form by the entrepreneur, it cannot explain transitions among legal forms. For instance, the analysis offers no insights into why and when a private corporation might go public.

A BRIEF LITERATURE OVERVIEW

The literature on financing the entrepreneurial firm is diverse. This is because the characteristics that define the entrepreneurial firm are multifold. Moreover, some of these characteristics are not unique to the entrepreneurial firm but hold for other companies as well. In the following we briefly describe six major economic problems associated with financing the entrepreneurial firm that have been dealt with in finance literature.

First, there is the problem of project choice or, synonymously, asset substitution. If the entrepreneur holds the equity and the outsider holds the debt, the insider has an incentive to choose excessively risky projects. This is because the debt holder shares the downside risk, but not the upside risk. By increasing project risk, the entrepreneur can shift part of the additional risk to the debt holder. The classic study in this area is Jensen and Meckling (1976). Note that the problem of asset substitution is not confined to the entrepreneurial firm. For instance, hired management of publicly traded firms might acquire stock or stock options in the firm as part of its compensation package.

Second, there is the problem of private control benefits. Management might take pleasure in being in control of the operations. The total value of the firm consists of its "social value," which is its fair market value, and the "private value," which is the value of the management's control benefits. The distinction between private and social values is important when it comes to the allocation of control rights between management and outside shareholders. Two seminal studies in this area are Grossman and Hart (1988) and Harris and Raviv (1988). For a textbook treatment, see Hart (1995). Note that private control benefits are not confined to the entrepreneur, but can apply to hired management as well.

Third, there is the problem of managerial entrenchment. Managerial entrenchment is not confined to the entrepreneurial firm, but might hold true for the firm with hired management as well. On one hand, the higher the fraction of equity owned by management, the closer the interests of management might be aligned to the interests of the outside shareholders. On the other hand, the more voting power management has, the more insulated it is from the disciplining forces of the market for corporate control. Entrenchment allows management to employ corporate resources for private benefits, for instance, through empire building or consumption of perquisites. Morck, Schleifer, and Vishny (1988) and McConnell and Servaes (1990) provide empirical evidence for managerial entrenchment. Zwiebel (1996) shows in a theoretical model how entrenched management can employ dividend payments as a means of committing to paying out free cash flow and keeping outside investors from intervening.

Fourth, free cash flow might be non-verifiable. When free cash flow is non-verifiable, the insider has the opportunity to divert it to consumptive use, for instance, by financing perquisites. Again, this problem is not confined to the entrepreneurial firm. On the other hand, non-verifiability of cash flows is seen as a significant problem in corporate finance and it earns considerable attention in finance literature. One thread of literature analyzes the role of debt as a way of forcing management to pay out free cash flow. Hart (1995) summarizes this literature in

a textbook treatment. Another thread of literature looks at the role of dividend payments as a device to discipline management by maintaining a sufficient level of debt. The aforementioned paper by Zwiebel (1996) fits into this category. Finally, a recent paper by Myers (2000) studies the role of dividend payments in an all-equity financed firm. He shows that dividend payments can be viewed as advance compensation on the outsider's equity investment. The dividend payment at the end of the fiscal year compensates the investor for the opportunity cost of capital that he incurs in the upcoming fiscal year.

Fifth, there is the problem of risk sharing. Entrepreneurs typically are risk-averse individuals who have most of their human capital and financial wealth invested in the firm. Outside equity is a means of sharing risk as it allows the entrepreneur to limit his exposure to the enterprise. For details, see a recent theoretical paper by Kirilenko (2001).

Sixth, there is the problem of asset complimentarity. In the entrepreneurial firm, management owns an asset that is complimentary to the firm's operating assets. This means that the entrepreneur's human capital and the operating assets are worth more in combination than in isolation. Complimentarity between the managerial asset and the operating assets is an attribute that is confined to the entrepreneurial firm. While hired management is dispensable, the entrepreneur is not. Myers (2000) briefly discusses this point.

A SIMPLE MODEL OF THE ENTREPRENEURIAL FIRM

In the following analysis, the entrepreneur is defined as a person who owns human capital that is complementary to the operating assets of a firm (a "project"). As mentioned above, assets are complementary if they are more valuable when used in combination than in isolation.[4] The entrepreneur can add value to the firm only by being in control of the day-to-day operations. He is indispensable for the operations because of his unique inventive and managerial skills. As opposed to a hired manager, the entrepreneur cannot be removed from the firm without hurting the project's net present value (NPV) of continuation. To keep matters simple, we assume that removing the entrepreneur from the firm reduces the NPV of continuation to zero or, put differently, lowers the going concern value to the level of the liquidation value.

The need for outside financing arises from the entrepreneur's wealth constraint. In terms of the model, the entrepreneur's initial wealth, w, falls short of the investment needed to start up the firm, $\bar{K}$, which we assume to be exogenous. The amount $\bar{K} - w$ needs to be financed by an (outside) investor. The liquidation value of the operating assets is observable, verifiable (at zero cost), and equal to $\bar{K}$ in every fiscal year. Because the operating assets do not depreciate, there is an infinite investment horizon. There is no uncertainty in the model.

The project generates perpetual free cash flow equal to

$$y_t = y = (1+\kappa)\cdot r \cdot \bar{K}$$

at the end of every fiscal year $t (t = 1,\ldots,\infty)$, where r is the marginal cost of capital (for the entrepreneur and the investor) and $\kappa > 0$ captures the value that the entrepreneur adds continually to the firm through his human capital. The condition $\kappa > 0$ implies that the project has positive NPV at the outset (i.e., the beginning of fiscal year 1), as well as positive present value (PV) of continuation at the end of every fiscal year. We assume that there are no (other) positive-NPV projects becoming available to the entrepreneur once the project has been started. Thus, there is no productive use for internally generated funds.

The fundamental problem of financing the entrepreneurial firm is the entrepreneur's ability to divert free cash flow. By virtue of being in control of the operations, the entrepreneur can route free cash flow into personal accounts (where it earns r at the margin). The amount of free cash flow that the entrepreneur diverts in fiscal year t is denoted z_t. Because all fiscal years are identical, we can write the following:

$$z_t = z \text{ for all } t,\ t = 1,\ldots,\infty.$$

We assume that while cash flows are observable (at zero cost) to outside investors, they are not verifiable (or verifiable only at prohibitively high costs). Were the project's cash flows verifiable (at zero cost), the problem of asymmetric information in financing the entrepreneurial firm would not exist. Because the entrepreneur would be able to commit to dividend payments, any positive-NPV project would be financed. However, with cash flows being non-

[4] See Hart (1995) for details on asset complementarity and its implications for firm organization.

verifiable, the entrepreneur's inability to commit to future dividends leads to underinvestment. We will show that for sufficiently low levels of initial wealth, w, or, equivalently, sufficiently low levels of project profitability, κ, there are positive-NPV projects that are not undertaken.

The investor can provide outside financing through debt or equity. In this article, only equity financing is considered.[5] The firm may be organized into a limited partnership or a corporation. The corporation may be private or public. Limited partnerships and private corporations tend to have few equity holders, with the entrepreneur being one of them. The public (traded) corporation is typically modeled in finance literature as a corporation with dispersed equity holders. The outside investor, in this case, is a multitude of small shareholders whose subjective probabilities of being pivotal to corporate decisionmaking may be viewed as zero. For simplicity, we assume that, in the limited partnership and the private corporation, the firm outsider is a single investor (or, equivalently, a group of block holders who act as a single investor).

In the next section, the problem of equity financing is analyzed for the limited partnership. In subsequent sections the firm is modeled as either a private or a publicly traded corporation. The analysis shows that the legal form is critical for the degree of bargaining power the parties have over the fiscal year's free cash flow, once the parties are invested. It turns out that this ex post bargaining power is greatest for the outside investor in the private corporation. Consequently, the private corporation faces the least restrictive outside equity constraint.

The Limited Partnership

The limited partnership consists of a general partner—the entrepreneur—and one or more limited partners, which are the outside investors. For simplicity we represent the firm outsiders by a single investor (or, equivalently, a small set of block holders that act as a single investor). Because there is no uncertainty in the model, the difference between the general and the limited partners lies solely in the decision rights over liquidation, which is the only dimension of project choice in this model. Generally, in limited partnerships the outside investor can withdraw his funds (at the end of the fiscal year), but cannot demand liquidation of (all or parts of) the operating assets.

It is immaterial to the mechanics of the model whether partial liquidation is an option or whether the project can be liquidated in full only. If partial liquidation is possible, the periodic free cash flow of the project, y, decreases by the percentage of the liquidated assets. The assumption of full liquidation, on the other hand, would mean that any liquidation, even if in part, reduces κ to zero, and with it the PV of continuation.[6]

The (outside) investor contributes the fraction x to the initial fixed investment, $\bar{K}$, while the entrepreneur contributes the fraction $1-x$:

$$x\,\bar{K} = \bar{K} - w \;\Leftrightarrow\; (1-x)\,\bar{K} = w\,.$$

We define $\tilde{x}$ to be the fraction of equity held by the investor. This fraction $\tilde{x}$ might be smaller than, equal to, or greater than the fraction of capital supplied by the investor, x, depending on whether the entrepreneur sells the equity at a premium, at par, or at a discount.

The partnership contract specifies that dividends are paid at the end of the fiscal year:

$$d_t = d = y - z \;\text{ for all }\; t,\; t = 1,\ldots,\infty\,.$$

Note that the fraction $1-\tilde{x}$ of the dividend payment goes to the firm insider.

Because all fiscal years are identical, we can write the PV of the future dividend stream at the beginning of any fiscal year as[7]:

$$\frac{y-z}{r} = \frac{d}{r}\,.$$

Investor's Continuation Constraint. The (outside) investor remains invested if the NPV of doing so is non-negative. In deriving this continuation constraint, we assume that the partnership contract allows the investor to withdraw his funds at the end of the fiscal year, which means that there is no vesting period. Upon demand, the entrepreneur has to pay to the investor the cash equivalent of the fraction $\tilde{x}$ of the liquidation value of the operating assets. Note that if the investor withdraws his funds, the entrepreneur has no incentive to pay dividends for the respective fiscal year. The investor remains invested in the firm for (at least) one more

[5] Hart (1995) presents an extensive treatment of debt financing of the entrepreneurial firm.

[6] The assumption of total liquidation has the advantage of being analogous to the assumption that a fixed amount of operating assets, $\bar{K}$, is needed to start the project.

[7] Note that $\frac{d}{r} \equiv \sum_{t=1}^{\infty} \frac{d}{(1+r)^t}$.

fiscal year if (and only if) the sum of this fiscal year's dividend payment and the PV of all future dividend payments is greater than or equal to the liquidation value. This yields the following continuation constraint for the firm outsider:

$$\tilde{x} \cdot (d + \frac{d}{r}) \geq \tilde{x}\, \overline{K} \;\Leftrightarrow\; d + \frac{d}{r} \geq \overline{K}.$$

In equilibrium, the investor's continuation constraint is satisfied at equality because the entrepreneur does not pay more in dividends than is needed to keep the outside equity capital in the firm:

$$\tilde{x} \cdot \overline{K} = \tilde{x} \cdot (d + \frac{d}{r}) \;\Leftrightarrow\; d = \frac{r\,\overline{K}}{1+r}.$$

Investor's Participation Constraint. The minimum level of dividend payments that satisfies the investor's continuation constraint need not fulfill the investor's participation constraint, i.e., his willingness to invest in the first place. For instance, if the outside equity were offered at par (which would imply $\tilde{x} = x$), the investor's participation constraint would be violated. This is because, for $\tilde{x} = x$, the investor's capital contribution, $x\overline{K}$, exceeds the PV of the dividend stream, xd/r:

$$x\,\overline{K} \;>\; x\,\frac{d}{r} = x\,\overline{K}\,\frac{1}{1+r}.$$

Because the entrepreneur cannot commit to dividend payments greater than what is needed to keep the outside equity capital in the firm, the investor is unwilling to finance the project in the first place. On the other hand, the entrepreneur can induce the investor to participate by issuing equity at a discount. Upon contributing $x\overline{K}$ in cash, the investor receives equity claims equal to $\tilde{x}\overline{K}$, $\tilde{x} > x$.

Note that, technically, the entrepreneur can issue equity at a discount by putting part of his own contribution, $w = (1-x)\overline{K}$, into equity reserves while selling the outside equity at face value.

The entrepreneur chooses the minimum $\tilde{x}$ that meets the investor's participation constraint:

$$r\,x\,\overline{K} = \tilde{x}\,d \quad \left[= \tilde{x}\;\frac{r\,\overline{K}}{1+r}\right].$$

This implies $\tilde{x} = x(1 + r)$ and consequently:

$$\tilde{x}\,d = x\,(1+r)\,d = x\,(1+r)\cdot\frac{r\,\overline{K}}{1+r} = x\,r\,\overline{K}.$$

By issuing equity at a discount, the entrepreneur transfers the following amount to the investor at inception of the project:

$$(\tilde{x} - x)\frac{d}{r} = (x\,[1+r] - x)\;\frac{d}{r} = r\,x\;\frac{d}{r} = x\,d.$$

Technically, the entrepreneur can do this by transferring an amount equivalent to one dividend payment, d, to the reserves. The implied transfer to the investor at the amount $x \cdot d$ increases the return on the outsider's investment, $x\overline{K}$, by an amount sufficient to fulfill his participation constraint. At the same time, the investor's continuation constraint remains satisfied at equality.

In the Appendix we show that, in competitive capital markets, the investor cannot do better than break even. This implies that the investor has no incentive to renegotiate the partnership contract at the end of the fiscal year. By renegotiating the contract, the investor cannot improve his position beyond the break-even point, which is what fulfills his participation constraint.

Entrepreneur's Participation Constraint. Because the (outside) investor breaks even, the entrepreneur's participation constraint is equivalent to the NPV decision rule:

$$\text{NPV} \geq 0 \;\Leftrightarrow\; \frac{y}{r} - \overline{K} \geq 0.$$

For $y/r \equiv z/r + d/r$ and $\overline{K} \equiv x\overline{K} + (1–x)\overline{K}$, we obtain

$$\text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + \frac{d}{r} - x\,\overline{K} - (1-x)\,\overline{K} \geq 0.$$

With $d/r \equiv (1–\tilde{x})d/r + \tilde{x}d/r$, we can write

$$\text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + (1-\tilde{x})\;\frac{d}{r} + \tilde{x}\frac{d}{r} - x\,\overline{K} - (1-x)\,\overline{K} \geq 0.$$

After rearranging terms we obtain

$$(1) \qquad \text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + (1-\tilde{x})\;\frac{d}{r} - (1-x)\,\overline{K} \geq 0.$$

The entrepreneur's participation constraint states—according to inequality (1)—that the PV of the diverted free cash flow, z/r, plus the received dividend payment, $(1–\tilde{x})d/r$, must not be lower than the initial cash contribution, $(1–x)\overline{K}$.

There is another way to read the entrepreneur's participation constraint. Adding and subtracting xd/r from inequality (1) yields

$$\text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + (1-x)\;\frac{d}{r} - (\tilde{x} - x)\;\frac{d}{r} - (1-x)\,\overline{K} \geq 0.$$

THE LIMITED PARTNERSHIP—A NUMERICAL EXAMPLE

Assume the following values for the exogenous variables in the limited partnership:

$$\overline{K} = 100;\ w = 70;\ r = 0.1;\ \kappa = 0.4\,.$$

Note that the project has positive NPV:

$$\frac{y}{r} - \overline{K} \equiv \frac{(1+\kappa)\, r\, \overline{K}}{r} - \overline{K} = 140 - 100 = 40 > 0\,.$$

If the entrepreneur invests all his initial wealth, the fraction of equity provided by the outsider reads

$$x \equiv \frac{\overline{K} - w}{\overline{K}} = 0.3\,.$$

Factoring in the equity discount, the outsider is assigned the following fraction of voting stock:

$$\tilde{x} \equiv (1+r)\, x = 0.33\,.$$

Dividend payments per fiscal year amount to

$$d \equiv \frac{r\, \overline{K}}{1+r} = \frac{0.1 \cdot 100}{1.1} = 9.\overline{09}\,.$$

The entrepreneur diverts the following amount into private accounts at the end of every fiscal year:

$$z \equiv y - d = 14 - \frac{10}{1.1} = 4.\overline{90}\,.$$

The outside equity constraint is fulfilled because the entrepreneur's initial wealth exceeds the equity discount granted the outsider:

$$0.7 = \frac{w}{\overline{K}} \;>\; \frac{r}{1+r} = 0.\overline{09}\,.$$

If (and only if) the entrepreneur's initial wealth amounts to not more than $9.\overline{09}$ percent of the amount needed to purchase the operating assets, the outside equity constraint is violated.

The investor breaks even. The outsider's return equals the opportunity cost of capital:

$$\frac{\tilde{x} \cdot d}{x \cdot \overline{K}} = \frac{0.33 \cdot 9.\overline{09}}{0.3 \cdot 100} = 10 \text{ percent}\,.$$

For $\tilde{x} \equiv x(1+r)$, we can write

$$\text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + (1-x)\,\frac{d}{r} - (x[1+r]-x)\,\frac{d}{r} - (1-x)\,\overline{K} \geq 0\,.$$

After rearranging terms, we obtain

$$(2)\quad \text{NPV} \geq 0 \;\Leftrightarrow\; \frac{z}{r} + (1-x)\,\frac{d}{r} - x\,d - (1-x)\,\overline{K} \geq 0\,.$$

Inequality (2) states that the PV of the diverted free cash flow, z/r, plus the dividend payment on the initial capital contribution, $(1-x)d/r$, must not fall short of the sum of the capital transfer to the investor, $(\tilde{x}-x)d/r \equiv xd$, and the initial capital contribution, $(1-x)\overline{K}$.

Outside Equity Constraint. Similar to the debt capacity constraint in debt financing, there is an outside financing constraint for equity. The constraint emerges from the entrepreneur's limited wealth. The more constrained the entrepreneur is, the higher the fraction of equity claims he has to grant the outsider. The outside equity constraint states that the fraction of equity claims held by the outsider, $\tilde{x}$, is limited to values less than one: $\tilde{x} < 1$. For $\tilde{x} \equiv x(1+r)$, this inequality results in:

$$x < \frac{1}{1+r}\,.$$

With x being equal to the fraction of equity injected by the outsider, we obtain

$$\frac{\overline{K} - w}{\overline{K}} < \frac{1}{1+r}\,.$$

After rearranging terms, the outside equity constraint reads

$$\frac{w}{\overline{K}} > \frac{r}{1+r}\,.$$

The inequality states that the fraction of outside equity in total liabilities must exceed a threshold that is solely determined by the marginal cost of (equity) capital. Because this condition is independent of the profitability of the project, κ, positive-NPV

projects will not get financed if the entrepreneur's initial wealth is sufficiently low.

Optimal Level of Outside Equity Financing. As shown, the outside equity constraint sets an upper limit to the fraction of outside equity in total equity. This does not imply that the entrepreneur finances the highest fraction possible through outside equity. In fact, the entrepreneur is indifferent between investing all his initial wealth, w, or investing any amount e, $0 < e < w$, that satisfies the outside equity constraint. As assumed, the entrepreneur and the investor face the same opportunity cost of capital. Because the investor breaks even, there is no difference in the costs of capital to the entrepreneur with respect to outside versus inside equity.

The Corporation

The corporation may be privately held (private corporation) or publicly traded (public corporation). We start out with the private type and maintain the assumption from the partnership model that the firm outsider is a single investor (or, equivalently, a small set of block holders that act as a single investor). In a subsequent section, we compare this organizational type to the public corporation, assuming that the outsider is a multitude of small shareholders.

There are important differences between limited partnerships and corporations with regard to the outside equity holder's control rights. First, while in the limited partnership the investor can withdraw his funds at the end of the fiscal year, in the corporation the investor can pull his funds only if he has command over the majority of votes that is necessary for liquidation.[8] This is because shareholders cannot sell in the aggregate, even though they can trade shares with each other. Consequently, in the corporation, minority investors are "locked in," which makes them vulnerable to opportunistic behavior by the entrepreneur. Second, unlike the limited partnership, the corporation enables the outside investor—if he has command over the necessary majority of votes—to remove the entrepreneur from the firm at the annual, end-of-fiscal-year shareholder meeting.

Note that, as in the partnership model, the optimal outcome demands the entrepreneur to be in control of the operations. Also, liquidation (in full or in part) is generally suboptimal because the going concern value always exceeds the liquidation value.

The distribution of voting rights influences the balance of power between insider and outsider when bargaining over the free cash flow at the end of the fiscal year. In the following we look at two cases of voting rights distribution, one in which the outsider has command over a simple majority of votes, and one in which the entrepreneur holds the majority. (We exclude the borderline case of each party holding 50 percent of the votes.) The outcome of this bargaining process determines the dividend payments, which in turn determines the firm's outside equity financing capacity.

Dividends and Allocation of Voting Rights. In the case where the investor holds the majority of votes, the outsider has the power to remove the entrepreneur from the firm or, equivalently, liquidate the firm.[9] On the other hand, the entrepreneur has the power to refuse to contribute his human capital to the firm henceforth. Any of these two non-cooperative actions by the investor or the entrepreneur would reduce the PV of next fiscal year's free cash flow from $(1+\kappa)r\bar{K}/(1+r)$ to $r\bar{K}/(1+r)$, with the fraction $\tilde{x}$ going to the investor and the fraction $1-\tilde{x}$ going to the entrepreneur. Moreover, we assume that, if the entrepreneur were removed (at the end of the fiscal year), the entrepreneur would have no incentive to pay dividends for the current fiscal year. Because cash flows are not verifiable, the entrepreneur would divert all of the fiscal year's free cash flow into personal accounts before leaving.

Nevertheless, the parties have an incentive to cooperate because cooperation generates a surplus that the parties can share. The cooperative outcome may be modeled as a Nash bargaining solution. In a Nash bargaining solution, each party is assigned the same degree of bargaining power, and thus they share the surplus from cooperation evenly. In fact, each party receives what it would receive if the parties did not cooperate plus half the gains from cooperation.[10]

The surplus from cooperation at the end of the fiscal year equals the PV of the entrepreneur's value added in the next fiscal year, $(\kappa r\bar{K})/(1+r)$. In sharing the surplus evenly, the entrepreneur pays out

[8] We assume that there is no supermajority rule in place, which means that all decisions are made with a simple majority.

[9] Removing the entrepreneur from the firm would result in a PV of continuation of zero, which is equivalent to liquidation.

[10] For details on the Nash bargaining solution, see Dagan, Volij, and Winter (2000). For a textbook example of a Nash bargaining solution in a corporate finance context, see Hart (1995).

$\kappa r\overline{K}/(2(1+r))$ to the investor, in addition to what the investor would receive in the noncooperative outcome, $\tilde{x}r\overline{K}/(1+r)$.[11] Consequently, the firm's total dividend payment equals

$$\frac{\kappa r\,\overline{K}}{1+r}+\frac{r\,\overline{K}}{1+r}=\frac{(1+\kappa)\,r\,\overline{K}}{1+r}.$$

Note that the dividend payment equals the PV of the free cash flow of the next fiscal year. This implies the following for the amount, z, that the entrepreneur diverts for personal use:

$$z=(1+\kappa)\,r\,\overline{K}-\frac{(1+\kappa)\,r\,\overline{K}}{1+r}$$

$$\Leftrightarrow\quad \frac{z}{r}=\frac{(1+\kappa)\,r\,\overline{K}}{1+r}.$$

In the case where the entrepreneur holds the majority of votes, the outsider has no bargaining power once he is invested. Because the investor has no ex post bargaining power, the entrepreneur has no incentive to pay dividends. Thus, an initial private equity offering in which less than 50 percent (plus one vote) of the shares are up for sale will fail. (The entrepreneur would have to offer the shares for free, which raises no funds.)

Investor's Participation Constraint. The (outside) investor expends the amount $x\overline{K}$ at the inception of the project and receives a dividend payment equal to $r\overline{K}(\kappa+2\tilde{x})/(2(1+r))$. Thus, the investor's participation constraint reads

$$r\,x\,\overline{K}\le\frac{r\,\overline{K}(\kappa+2\tilde{x})}{2\,(1+r)}.$$

Let $\delta/(1+\delta)$, $\delta>-1$, be the percentage discount (or premium, if negative) at which the entrepreneur issues the equity to the outsider; then we can write the investor's participation constraint as follows:

$$r\,x\,\overline{K}\le\frac{r\,\overline{K}(\kappa+2x\,[1+\delta])}{2\,(1+r)}.$$

Rearranging terms yields

$$\delta\ge r-\frac{\kappa}{2x}.$$

The entrepreneur's participation constraint is subject to $1>\tilde{x}\equiv x(1+\delta)>1/2$, which demands that the outside investor holds more than 50 percent of the voting stock. Remember that the entrepreneur chooses the discount (premium) such that the investor merely breaks even or, in other words, such that the outsider's participation constraint is satisfied at equality.

Because the fraction of equity held by the entrepreneur, $1-\tilde{x}$, can be zero, there is no upper bound on δ; this means that, for any $1\ge\tilde{x}\equiv x(1+\delta)>1/2$, there is a value $\delta>-1$ at which the investor breaks even. We assume that the entrepreneur can divert any excess amount raised in the private equity offering, which is analogous to our assumption that the entrepreneur can divert free cash flow.

Note that there is a lower bound on the equity discount: $\delta>-1$. The fraction of capital injected by the investor, x, must be positive and so must be the fraction of (voting) equity that emanates from this initial capital contribution, $\tilde{x}$.

Entrepreneur's Participation Constraint. The entrepreneur expends the amount $(1-x)\overline{K}$ at inception of the project, receives periodic dividend payments equal to

$$\frac{(1-\tilde{x})\,r\,\overline{K}}{1+r}+\frac{\kappa r\,\overline{K}}{2\,(1+r)},$$

and can divert the amount $(1+\kappa)r^2\overline{K}/(1+r)$ into personal accounts at the end of every fiscal year. Thus, the entrepreneur's participation constraint reads

$$\frac{(1+\kappa)\,r\,\overline{K}}{1+r}+\frac{(1-\tilde{x})\,\overline{K}}{1+r}-\frac{\kappa\,\overline{K}}{2\,(1+r)}+(1-x)\,\overline{K}\;\ge\;0.$$

Substituting $x(1+\delta)$ for $\tilde{x}$ yields

$$\delta\le\frac{\kappa\,(r+\frac{1}{2})}{x}+r.$$

As with the investor's participation constraint discussed in the preceding section, the entrepreneur's participation constraint is subject to $1>\tilde{x}\equiv x(1+\delta)>1/2$, which demands that the outside investor holds more than 50 percent of the voting stock.

Outside Equity Constraint. The entrepreneur's limited wealth imposes an outside equity constraint on the firm. It is this constraint that necessitates outside equity and—at the same time—may impose an upper bound on the fraction of outside equity in total liabilities. The outside equity constraint states that the fraction of equity claims held by the outsider, $\tilde{x}$, must not exceed unity: $\tilde{x}\le 1$. For $\tilde{x}\equiv x(1+\delta)$, this inequality results in

[11] Note that this outcome fulfills the investor's continuation constraint by exceeding the opportunity cost of capital.

$$x \leq \frac{1}{1+\delta}.$$

From the discussion of the investor's participation constraint, we know that the outsider breaks even, which implies the following equality for the percentage by which the fraction of equity claims held by the outsider, $\tilde{x}$, exceeds the fraction of the equity capital contributed by the outsider, x:

$$\delta = r - \frac{\kappa}{2x}.$$

Inserting the right-hand-side expression into the above inequality yields

$$x \leq \frac{1}{1+r-\frac{\kappa}{2x}}.$$

With x being equal to the fraction of capital injected by the outsider, $(\overline{K}-w)/\overline{K}$, we obtain

$$\frac{\overline{K}-w}{\overline{K}} \leq \frac{1+\frac{\kappa}{2}}{1+r}.$$

After rearranging terms, the outside equity constraint reads

$$\frac{w}{\overline{K}} \geq \frac{r-\frac{\kappa}{2}}{1+r}.$$

Except in the case of the limited partnership, the minimum fraction of initial wealth, w, in total assets, $\overline{K}$, that the entrepreneur needs to launch the project depends not only on the opportunity cost of capital, r, but also on the project's profitability, κ.

Limited Partnership Versus Private Corporation

We use the model to explain the choice between the limited partnership and the (private) corporation. We can show that, in the private corporation, the outside equity constraint is less likely to bind. For low levels of initial wealth, the entrepreneur may be unable to attract outside equity when organizing the project into a limited partnership. The private corporation, on the other hand, lends greater ex post bargaining power to the outsider, which increases the firm's outside equity capacity. Note that the model is not dynamic, which implies that it cannot explain a transition from the partnership to the corporation, or vice versa.

Assuming competitive capital markets, the outside investor cannot do better than break even. The entrepreneur anticipates the investor's ex post bargaining power and issues the equity at the lowest discount (or highest premium) that fulfills the investor's participation constraint. Thus, if the project gets financed, the entrepreneur reaps all the NPV. Also, if the project gets financed, the NPV is independent of the choice of the legal form. Consequently, the entrepreneur is indifferent about the legal form if (and only if) this decision is inconsequential to whether the project gets financed.

The only reason why a project of positive NPV might not get financed is the entrepreneur's insufficient initial wealth, w. In the limited partnership, the outside equity constraint (subject to the investor's participation constraint being satisfied) reads

$$\frac{w}{\overline{K}} > \frac{r}{1+r}. \tag{3}$$

Inequality (3) states that the entrepreneur has to finance more than the fraction $w/\overline{K}$ of the project's initial expenses, $\overline{K}$, through initial wealth. The threshold level $r/(r+1)$ is solely determined by the opportunity costs of capital, r.

In the (private) corporation, the outside equity constraint is given by

$$\frac{w}{\overline{K}} \geq \frac{r-\frac{\kappa}{2}}{1+r}. \tag{4}$$

Inequality (4) demands that the ratio of the entrepreneur's initial wealth, w, to the initial expenses needed to launch the project, $\overline{K}$, must meet a threshold level that is determined by the opportunity cost of capital, r, and the profitability of the project, κ.

Comparing the constraints across legal forms shows that the outside equity constraint is more restrictive in the limited partnership model than in the private corporation. This is because, in the private corporation, the ex post bargaining power of the investor at the end of every fiscal year allows him to capture—in a Nash bargaining solution—half the value added of the upcoming fiscal year. For a sufficiently profitable project, i.e., $\kappa \geq 2r$, the outside equity constraint never binds in the corporation model. Remember that the outside equity constraint in the limited partnership model is independent of the project's profitability.

Going Public

We now analyze the public corporation. In finance literature, the term public corporation

THE PRIVATE CORPORATION—FOUR NUMERICAL EXAMPLES

We assume the same set of values for the exogenous variables of the private corporation as we did in the numerical example for the limited partnership:

$$\bar{K} = 100;\ w = 70;\ r = 0.1;\ \kappa = 0.4\,.$$

As established for the limited partnership, the project has positive NPV. This is because the project returns exceed the opportunity cost of capital by the factor $\kappa = 40$ percent. As in the partnership, the outsider has to contribute at least the fraction $x \equiv (\bar{K} - w)/\bar{K} = 0.3$ to the firm's equity capital.

An important difference of the private corporation to the limited partnership is the constraint $\tilde{x} \equiv (1+\delta)x > 1/2$, which demands that the outsider holds the majority of votes. This condition restricts the entrepreneur in his choice of x (the fraction of capital contributed by the outsider) and $\delta/(1+\delta)$ (the equity discount). For instance, assume that the entrepreneur contributes $1-x = 0.7$ to the firm's capital, just as in the numerical example for the limited partnership. The value of δ at which the outsider breaks even would then read

$$\delta = r - \frac{\kappa}{2x} = -0.5\overline{6}\,.$$

This would mean that the outside equity is issued at a discount equal to $\delta/(1+\delta) \approx -1.31$, or in other words, at a premium of about 131 percent. At this premium, though, the outside investor would not have command over the majority of votes:

$$0.13 = x \cdot (1+\delta) \equiv \tilde{x} < \frac{1}{2}\,.$$

Consequently, the entrepreneur chooses to contribute a lower fraction to the firm's equity than what he is able to. For instance, the entrepreneur might contribute the fraction $1-x = 0.1$ only.

For $1-x = 0.1$, the value of δ at which the outsider breaks even, reads

$$\delta = r - \frac{\kappa}{2x} = -0.1\overline{2}\,.$$

This would mean that the outside equity is issued at a discount equal to $\delta/(1+\delta) \approx -0.14$, or, in other words, at a premium of approximately 14 percent. The outsider would end up with $\tilde{x} \equiv (1+\delta)\cdot x = 79$ percent of the voting stock.

The dividend payment amounts to

$$d \equiv \frac{(1+\kappa)r\bar{K}}{1+r} = \frac{1.4 \cdot 0.1 \cdot 100}{1.1} = 12.\overline{72}\,.$$

The amount that the entrepreneur diverts into personal accounts equals

$$z \equiv y - d = 14 - 12.\overline{72} = 1.\overline{27}\,.$$

The outside investor breaks even:

$$9 = 0.1 \cdot 0.9 \cdot 100 = rx\bar{K}$$
$$= \frac{r\bar{K}(\kappa + 2x[1+\delta])}{2(1+r)}$$
$$= \frac{0.1 \cdot 100 \cdot (0.4 + 2 \cdot 0.9 \cdot [1 - 0.1\overline{2}])}{2 \cdot 1.1} = 9\,.$$

The outside equity constraint is fulfilled:

$$0.3 = \frac{w}{\bar{K}} \geq \frac{r - \frac{\kappa}{2}}{1+r} = -0.0\overline{9}\,.$$

Because the project is sufficiently profitable, the outside equity constraint never binds. For $\kappa \geq 2r$ the entrepreneur is able to attract outside equity without contributing his own capital. This is in sharp contrast to the numerical example of the limited partnership, where the entrepreneur has to contribute more than $9.\overline{09}$ percent of the initial capital expenditure.

The outside equity constraint is relevant only if the project is not sufficiently profitable, i.e., if $\kappa < 2r$. Even in this case, the outside equity constraint is less restrictive than in the limited partnership. For instance, in the partnership the project cannot be financed if the entrepreneur's wealth, w, amounts to less than $9.\overline{09}$. In the private corporation, though, the entrepreneur can attract outside equity for a project with low profitability $\kappa = 0.1$ even if his wealth is as low as $w = 4.\overline{54}$. The entrepreneur would issue the equity at a discount of $4.\overline{54}$, granting the outsider $\tilde{x} = 100$ of the voting capital.

represents listed companies with dispersed shareholder structures.

Dispersed shareholders face high costs of collective action. These costs restrain the bargaining power the shareholders have once they are invested. We assume that if (and only if) the dividend payments fall short of the opportunity cost of capital, the shareholders remove the entrepreneur from the corporation or, equivalently, liquidate the project. Note that without this assumption, the entrepreneur pays no dividends.

Investor's Continuation Constraint. As with the private corporation, outside equity financing is possible only if the outside shareholders have command over a majority of votes: $\tilde{x} > 1/2$. Without a majority of votes, the investor cannot liquidate, and, consequently, the entrepreneur has no incentive to pay dividends.

In the event of liquidation, the shareholders can withdraw their fraction of assets. Thus, the continuation constraint reads

$$\tilde{x}\, d \geq \tilde{x}\, \frac{r\,\overline{K}}{1+r}\,.$$

This continuation constraint is equivalent to the continuation constraint in the partnership model.

Investor's Participation Constraint. Subject to $\tilde{x} > 1/2$, the (outside) investor's participation constraint reads

$$r\, x\, \overline{K} \;\leq\; \tilde{x}\, d \left[= \tilde{x}\, \frac{r\,\overline{K}}{1+r}\right].$$

Except for the condition $\tilde{x} > 1/2$, this participation constraint is identical to the investor's participation constraint in the partnership model.

Entrepreneur's Participation Constraint. Under the assumption of competitive capital markets, the investor breaks even, which implies that the investor's participation constraint is fulfilled at equality. For the investor to break even, the entrepreneur has to issue the outside equity at a discount:

$$\tilde{x} = x\,(1+r)\,.$$

Because the investor does not do better than break even, the entrepreneur reaps the whole (positive) NPV, which implies that the entrepreneur's participation constraint is satisfied.

Equivalence of Legal Forms. In summary, the public corporation model is equivalent to the partnership model when the constraint $\tilde{x} > 1/2$ is added and the outside equity constraint is allowed to hold at equality:

Table 1

Outside Equity Constraints

Legal form	Outside equity constraint
Limited partnership	$\frac{w}{\overline{K}} > \frac{r}{1+r}$
Private corporation	$\frac{w}{\overline{K}} \geq \frac{r-\frac{\kappa}{2}}{1+r}$
Public corporation	$\frac{w}{\overline{K}} > \frac{r}{1+r}$
w	Entrepreneur's initial wealth
$\overline{K}$	Expenses needed to purchase the operating assets
r	Marginal cost of capital
κ	Project profitability (percentage by which the project return exceeds the opportunity cost of capital)

$$\frac{w}{\overline{K}} \;\geq\; \frac{r}{1+r}\,.$$

In contrast to the limited partnership, the wealth of the entrepreneur in the corporation may adopt a value of zero.

For the constraint $\tilde{x} > 1/2$, which demands the outsider to hold the majority of votes, we obtain the following after rearranging terms:

$$\frac{w}{\overline{K}} \;<\; \frac{\frac{1}{2}+r}{1+r}\,.$$

This constraint never binds, because the entrepreneur can choose to invest only the amount $e \leq w$ of his wealth.

The outside equity constraints of all three legal forms are summarized in Table 1.

CONCLUSION

In this article we analyzed outside equity financing of the entrepreneurial firm when cash flows are non-verifiable. The study covered three legal forms: the limited partnership, the private corporation, and the public corporation. We showed that the limited partnership and the public corporation are equivalent with regard to the outside investor's

bargaining power once he is invested. Of the three analyzed legal forms, the private corporation renders the most ex post bargaining power to the outside investor. Consequently, when organizing the project into a private corporation, the entrepreneur faces the least restrictive outside equity constraint. This is because the outside investors in the private corporation are block holders whose costs of collective action are low. The block holders have the power to remove the entrepreneur from his post (or, equivalently, liquidate the firm). Clearly, this presupposes that the outsiders hold the majority of votes. If outside equity holds a minority interest in the corporation, the insider has no incentive to pay dividends and will divert all free cash flow for personal use.

The article underlines the importance of venture capital financing. Venture capital allows positive-NPV projects that are organized into private corporations to be financed even when the entrepreneur is strongly wealth-constrained. For less wealth-constrained entrepreneurs, the three legal types are equivalent. If the entrepreneur's wealth is sufficiently low, however, positive-NPV projects cannot be financed in either legal form. This is because of the entrepreneur's inability to commit to a level of dividend payments that allows the outside investor to break even.

REFERENCES

Dagan, Nir; Volij, Oscar and Winter, Eyal. "A Characterization of the Nash Bargaining Solution." Unpublished manuscript, Brown University, Washington University, and Hebrew University, 2000, < http://www.nirdagan.com/research/ > .

Grossman, Sanford J. and Hart, Oliver D. "One Share–One Vote and the Market for Corporate Control." *Journal of Financial Economics*, January/March 1988, *20*(1/2), pp. 175-202.

Harris, Milton and Raviv, Artur. "Corporate Governance: Voting Rights and Majority Rules." *Journal of Financial Economics*, January/March 1988, *20*(1/2), pp. 203-35.

Hart, Oliver. *Firms, Contracts, and Financial Structure*. New York: Clarendon Press, 1995.

Jensen, Michael C. *A Theory of the Firm: Governance, Residual Claims, and Organizational Forms*. Cambridge, MA: Harvard University Press, 2000.

_____ and Meckling, William H. "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure." *Journal of Financial Economics*, October 1976, *3*(4), pp. 305-60.

Kirilenko, Andrei A. "Valuation and Control in Venture Finance." *Journal of Finance*, April 2001, *56*(2), pp. 565-87.

McConnell, John J. and Servaes, Henri. "Additional Evidence on Equity Ownership and Corporate Value." *Journal of Financial Economics*, October 1990, *27*(2), pp. 595-612.

Milgrom, Paul and Roberts, John. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

Morck, Randall; Shleifer, Andrei and Vishny, Robert. "Management Ownership and Market Valuation: An Empirical Analysis." *Journal of Financial Economics*, January/March 1988, *20*(1/2), pp. 293-315.

Myers, Stewart C. "Outside Equity." *Journal of Finance*, June 2000, *55*(3), pp. 1005-37.

Zingales, Luigi. "In Search of New Foundations." *Journal of Finance*, August 2000, *55*(4), pp. 1623-53.

Zwiebel, Jeffrey. "Dynamic Capital Structure Under Managerial Entrenchment." *American Economic Review*, December 1996, *86*(5), pp. 1197-215.

Appendix

Is the partnership contract renegotiation-proof, or does the investor have an incentive to demand renegotiation of the contract by threatening to withdraw his funds?

If at the end of the first fiscal year the investor threatens to withdraw his funds, the entrepreneur has the option of not making the last dividend payment and substituting (renegotiation-free) debt for equity by issuing an infinitely lived bond, a console. The console would have a par value equal to $\tilde{x}\bar{K}$ and sell for $x\bar{K}$. The periodic, annual interest payment would amount to $\tilde{x}d$. If the entrepreneur defaults on servicing the console, the assets transfer to the debt holder either entirely (total liquidation, undertaken by the investor) or partially (partial liquidation, undertaken by the entrepreneur). Because the operating assets are worth more in the project than outside, the entrepreneur has an incentive to make the agreed-upon payments. On the other hand, the entrepreneur's ability to substitute a console for outside equity implies that the limited partner has no bargaining power when trying to renegotiate the partnership contract. Thus the partnership contract is renegotiation-free, which in turn implies that outside equity is equivalent to debt.

Dollarization as a Monetary Arrangement for Emerging Market Economies

Gaetano Antinolfi and Todd Keister

Ecuador and El Salvador have recently adopted the U.S. dollar as legal tender, replacing their own national currencies.[1] This same move has received serious attention in policy debates in both Argentina and Mexico. Abandoning the national currency is a decision with far-reaching economic and political implications that are not well understood. In response to this phenomenon, a growing literature has aimed at evaluating the economic costs and benefits of "dollarizing." In this article, we provide an overview of the emerging literature and point out some issues that we feel warrant further research.[2]

Throughout, we focus on *official dollarization*, where the U.S. dollar (or some other currency) replaces the national currency as legal tender. Unofficial dollarization, where private agents use a foreign currency as a substitute for the domestic currency, is already widespread in Latin America and elsewhere. We focus on Latin America and the U.S. dollar because of the recent events and policy debates mentioned above. Most of the issues we discuss, however, would apply to any country considering the official adoption of a foreign currency.

Discussions of the optimal monetary and exchange rate arrangements for an emerging market economy have traditionally centered on fixed or flexible exchange rates or (most often) some hybrid of the two, perhaps combined with capital controls or other regulatory measures. We begin our discussion by examining the causes of the current surge of interest in official dollarization. We then turn to the details of the issues that we feel are most important in analyzing the potential costs and benefits of dollarizing.

Gaetano Antinolfi is an assistant professor of economics in the Department of Economics, Washington University, and a visiting scholar at the Federal Reserve Bank of St. Louis. Todd Keister is an assistant professor of economics in the Departamento de Economía and Centro de Investigación Económica, Instituto Technológico Autónomo de México (ITAM) and a visiting assistant professor in the Department of Economics, University of Texas at Austin. The authors would like to thank James Morley, Patricia Pollard, Bob Rasche, Mika Saito, and Frank Schmid for helpful comments. William Bock provided research assistance.

WHY CONSIDER DOLLARIZATION?

Financial Crises

The current interest in official dollarization is largely a reaction to the recent string of currency crises. In the past decade, these crises have affected numerous countries, both industrialized (Italy and the United Kingdom in 1992) and emerging markets (Mexico in 1994, and East Asia and Brazil in 1997). Comparing the crises in industrialized countries with those in emerging markets reveals an important difference: although these crises are not costly in terms of lost output for industrialized economies, they are extremely costly for emerging market economies.[3] For example, in 1995 Mexican gross domestic product (GDP) declined by 7 *percent* in real per capita terms. (In the years before the crisis, for comparison, real per capita growth ranged between 3 percent and 10 percent.) Moreover, when one emerging economy suffers a crisis, others are often hit by interest rate increases and a recession, as happened in Argentina following the Mexican crisis. This phenomenon is known as *contagion*.

The events in emerging market economies share certain characteristics that allow us to identify a typical "anatomy" of a crisis.[4] Beforehand, there is an incipient capital inflow and a corresponding current account deficit. The onset of the crisis is marked by a sudden capital outflow and a large devaluation of the exchange rate. There is often a crisis in the banking system at about the same time.[5] The result is a sharp and painful fall in output. Much of the current interest in dollarization stems from a strong desire to avoid such crises in the future. Before discussing the potential costs and benefits of

[1] Guatemala has also recently adopted the U.S. dollar as legal tender, but it has decided to maintain its own currency in circulation, without fixing a parity with the dollar.

[2] The interested reader can find a good, basic introduction to the topic of dollarization in Chang (2000).

[3] This reflects the general finding that an exchange rate devaluation is usually contractionary for emerging markets, whereas it is typically expansionary for industrialized countries. See, for example, Edwards (1989).

[4] For a detailed discussion, see Calvo (2000).

[5] Kaminsky and Reinhart (1999) empirically show that banking crises tend to precede exchange rate crises.

dollarizing, we look at some of the more traditional approaches to these problems and why they seem to be falling out of favor.

The Fear of Floating

One approach that naturally comes to mind (to an economist, at least) is to allow prices and quantities to be determined by supply and demand in markets. The definition of a flexible exchange rate system is exactly this: the price of one currency relative to another is determined by the market without any intervention by central banks. That is to say, any current account deficit has to be financed entirely by capital inflows (a financial account surplus) and vice versa, without any change in official reserves.

In reality, however, we do not observe many countries with truly flexible exchange rate systems. Rapid growth in world capital markets has led to a substantial increase in the size of international capital flows. At times, these flows become very volatile; indeed, as we mentioned above, a sudden reversal in capital flows is the typical "spark" of a crisis. Under a pure flexible exchange rate system, such volatility in capital flows causes corresponding volatility in the exchange rate. A volatile exchange rate, in turn, means that relative prices in the economy are volatile, which can be very disruptive to real economic activity.

Calvo and Reinhart (2000) have termed the unwillingness to let exchange rates be completely determined in markets "the fear of floating." They also point out several additional reasons why emerging market economies seem to be averse to floating exchange rates. These include high levels of dollar-denominated debt, high-exchange-rate pass through (reflected in domestic inflation), and in general an adverse effect of currency instability on credit market access. In support of their argument, they conduct an empirical analysis comparing the announced exchange rate regime of countries to the actual exchange rate behavior. Their findings indicate that countries classified as letting their exchange rate float, in general, do not. Hence it seems that very few, if any, countries are willing to take this approach.

The Costs of Capital Controls

Sudden reversals in the flow of capital have been an important and particularly damaging aspect of currency crises. If capital market volatility is the problem, one way of avoiding it is to introduce capital controls. Clearly the aim of such a policy would not be to stop capital inflows, because emerging market economies rely on them for investment, but to diminish their volatility. There is evidence indicating that capital controls involving taxes and reserve requirements can change the composition of capital inflows in favor of long-term investment, and thereby decrease the likelihood of large, sudden outflows. Calvo and Reinhart (1999), however, caution that these results may depend on the accounting classifications of capital flows. In addition, Edwards (1999) argues that, when analyzing the maturity of a country's foreign debt, the relevant concept is residual maturity[6] rather than contractual maturity. Using data from Chile, Edwards shows that short-term capital controls had a limited effect on Chile's residual maturity of foreign debt and that Chile had higher residual maturity than Mexico (a country without capital controls) at the end of 1996.

More generally, capital controls are typically not considered sound economic policy because they limit the ability of a country to borrow and invest, they hinder international risk sharing and technology transfer, and they prolong the survival of unsustainable domestic policies. The main practical objection to capital controls, however, is that they create a strong incentive for tax evasion and require a costly enforcement apparatus. These problems make them poor candidates for permanent solutions.[7]

The Vanishing Intermediate Regime

The unwillingness to let exchange rates float and to use direct capital controls has pushed countries toward "intermediate" exchange rate regimes in which official intervention is used to keep the exchange rate within predetermined bounds. This move, however, has been accompanied by the recent crises mentioned previously. This association has led many observers to claim that intermediate exchange rate regimes are no longer viable for emerging market economies. These observers claim that only extreme (totally fixed or totally flexible) exchange rate regimes are viable for emerging market economies. Eichengreen (forthcoming) colorfully likens

[6] Residual maturity is measured by the value of a country's liabilities that are held by foreigners and mature within a year.

[7] See De Grauwe (1996, Chapter 11) and Neely (1999) for an extensive assessment of capital controls. Calvo and Reinhart (1999) provide a discussion related to the context of dollarization.

adopting an intermediate regime to "painting a bull's eye on the forehead of the central bank governor and telling speculators to 'shoot here'."[8] There are, of course, situations for which some authors are willing to defend intermediate regimes as appropriate, but they are generally viewed as temporary remedies.[9] Fischer (2001) presents empirical evidence that the proportion of emerging market economies using intermediate regimes has indeed declined over the past decade.

An important question is how extreme a policy must be in order to avoid the problems associated with the middle ground. Even a currency board has proven not to be extreme enough in some ways. Under this arrangement, the central bank commits to back its monetary base entirely with foreign reserves at all times; thus, a unit of domestic currency can be introduced into the economy only if an equivalent amount of foreign reserves is obtained. In principle, this system is equivalent to dollarization. However, even though Argentina has been operating under a currency board since 1991, the interest rate differentials between peso-denominated and dollar-denominated debt remain and have widened during periods of financial turmoil, as with the Brazilian and Mexican crises (see Figure 1). This indicates that financial markets believe there to be a significant probability that the currency board will be abandoned under such circumstances, and the Argentine economy has suffered as a result.

THE KEY ISSUES

We now turn to what we see as the key issues in evaluating the costs and benefits of dollarization. Two of the primary benefits of dollarization are straightforward: exchange rate volatility (against the dollar) and exchange rate crises would be eliminated, and in most cases the inflation rate would be lowered substantially. One of the costs is also fairly straightforward, although occasionally misunderstood: the loss of seignorage revenue. We begin our discussion with this issue.

There are other costs and benefits that are more subtle and difficult to measure. Dollarization implies the loss of monetary policy, but, if it enhances the credibility of economic policy, dollarization could lower interest rates and substantially decrease the likelihood of future financial crises. If it increases economic integration with the United States, dollarization could yield substantial benefits in both product and financial markets. An important concern, however, is that dollarization would limit the ability of the central bank to act as a lender of last resort. We discuss these issues in turn below. Finally, a discussion of dollarization would not be complete without looking at the "initial conditions" in which many emerging market economies currently find themselves and approaching the issue from the perspective of the United States.

Seignorage Revenue

An obvious cost of dollarization is the loss of the seignorage revenue that comes with the power to print fiat currency. The size of the flow of seignorage revenue depends on both the rate of growth of output and the rate of inflation. For some emerging market economies, it constitutes a substantial fraction of government revenues. With any other fixed-exchange-rate arrangement, seignorage revenues are present in some form. In particular, under a currency board, newly printed domestic money is used to buy interest-bearing foreign reserves. Dollarization entails losing this interest. It also entails buying back the domestic monetary base using foreign reserves and, therefore, losing the interest on this stock of reserves as well. Velde and Veracierto (1999) calculate this latter number for Argentina to be $658 million, or 0.2 percent of GDP, per year.

Note that computing the present level of seignorage revenue and calling that a "cost" of dollarization is clearly a mistake in most cases. Chang and Velasco (2000b) make this point: If a country dollarizes in order to lower its inflation rate, this reflects a decision that the benefits of lower inflation outweigh the value of the revenue that higher inflation brings. Instead, one should focus on the seignorage revenue that would have been earned at the new, lower inflation rate. This is the "loss" in seignorage revenue relative to the (ideal) case where the inflation rate is (somehow) lowered without dollarizing.

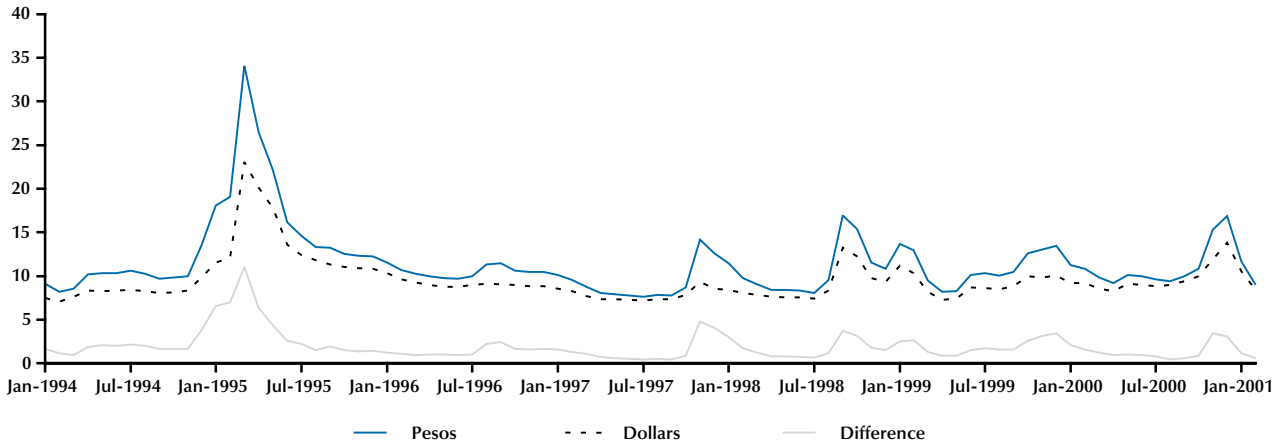
There are two reasons to believe that this amount may still overestimate the true revenue loss from dollarizing. First, a dollarizing country may be able to negotiate a deal with the United States under which it receives some of the increased U.S. seignorage revenue (which could equal the "loss" calcu-

[8] See also Obstfeld and Rogoff (1995) and Summers (1999; 2000, p.8). For a classification of different exchange rate regimes, see Frankel (1999).

[9] See Mussa et al. (2000), who argue that an unsustainable policy need not be undesirable in the short run, and Frankel (1999), from whom we borrowed the title of this section.

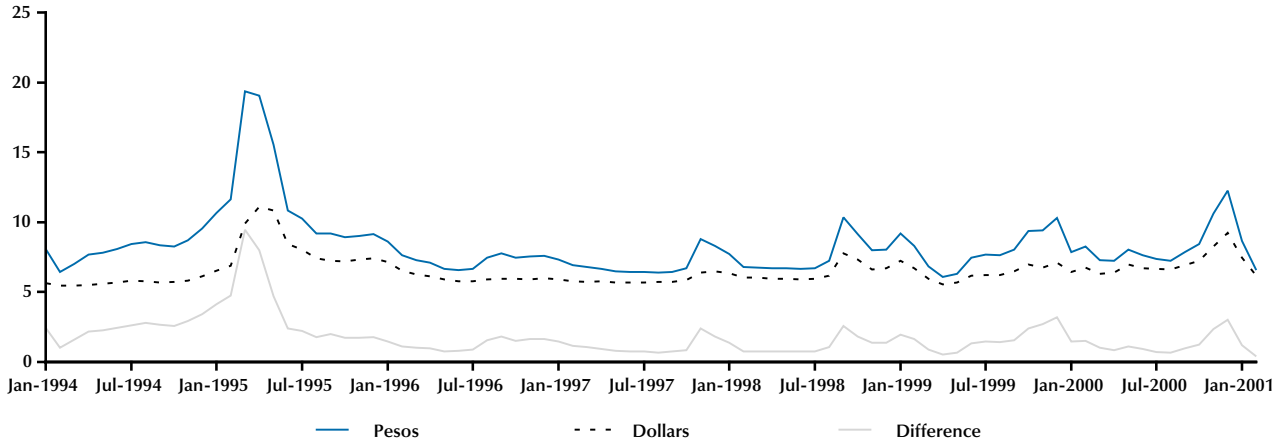
Figure 1

Interest Rates on Loans Given to Prime Companies



SOURCE: Central Bank of Argentina.

Interest Rates on Deposits (30- to 59-Day Term)



SOURCE: Central Bank of Argentina.

lated above).[10] Second, a large part of the reason for dollarizing is to create a more stable economic environment that will encourage investment and growth. While it is extremely difficult to make quantitative predictions about the size of this effect, it is clear that the increase in tax revenue from increased economic activity should at least partially offset the loss of seignorage revenue.

Regarding this last point, however, it is important to note that the increase in tax revenue would take time to develop. In the meantime, a government with lower revenues would have to decrease expenditures, increase taxes, or increase the public debt. To the extent that the loss of seignorage revenue is compensated by an increase in government borrowing, it may not be the case that a stable currency necessarily provides more macroeconomic stability. This is an indication that the fiscal plan accompanying a dollarization would be critical to its success.

[10] Such a plan was actually proposed as part of the International Monetary Stability Act, introduced by then Senator Connie Mack of Florida. Details can be found in Mack (2000).

Fiscal Consequences

Because dollarization entails a loss of both seignorage revenue and independent monetary policy, it is likely to have important consequences for the conduct of fiscal policy. Sims (2001) argues against dollarization for precisely this reason. He argues that the option value of issuing fiat debt (which can be defaulted on through inflation) is too high to surrender because inflation is part of an optimal taxation scheme. In support of his argument, Sims computes the unexpected component of U.S. government debt yields and shows that it is substantial. His calculations show that fiat debt worked to relax the government budget constraint in times of high economic turmoil (such as the oil crisis of 1973). One possible interpretation of this fact is that, without fluctuations in the unexpected component of government bond yields, more variability would have been observed in taxation and government expenditure, which may have been very costly. More research (as Sims acknowledges) is needed to evaluate (i) how much of the variability in the unexpected component of government bond yields actually reflects inefficient variation in monetary and fiscal policy that is better avoided and (ii) how much of it reflects an "optimal" response to real shocks.

Similarly, Chang and Velasco (2000b) argue that an optimal taxation plan would always entail surprise inflation (or devaluation) because this acts as a lump-sum tax and therefore is non-distortionary. Dollarization removes the ability of the government to use this tax. The contributions of Sims (2000) and Chang and Velasco (2000b) in this way stress the potentially high costs of losing flexibility in economic policy. Surprise inflations, however, cannot be repeatedly engineered, and anticipated inflation is typically *not* part of an optimal taxation plan. Hence the government has a time-consistency problem; it wants to convince people that it will not engineer an inflation increase, but once people are convinced, it wants to surprise them. Because people know this, the economy can end up in a situation of anticipated inflation. If this problem is very costly to the economy, then the benefit of using dollarization to solve it may easily outweigh the cost of the lost flexibility.[11]

Chang and Velasco (2000b) go on to point out that dollarization might decrease the incentives for fiscal discipline. Lack of fiscal discipline, in turn, may mean that crises due to high sovereign default risk would persist and the economy would not benefit from lower interest rates. Would the adoption of the dollar imply more or less fiscal discipline? Chang (2000) argues that, under alternative arrangements, changes in exchange rates or interest rates make the costs of a lack of fiscal discipline immediate. Dollarization would take those incentives away by allowing the costs of present fiscal looseness to be shifted to the future (in terms of higher future taxes, for example). Hence the incentives for fiscal discipline would decrease.

Although these incentive problems are real, it is important to recognize that there are other factors working in the opposite direction. First, as noted above, emerging markets depend heavily on foreign capital, and capital outflows could serve to make the cost of a lack of fiscal discipline immediate. Second, the incentive for domestic investors to monitor and put political pressure on the government for fiscal discipline would be higher. Heavy government borrowing would be perceived to induce macroeconomic instability and would cause interest rates for all domestic borrowers to increase. Through these channels, market discipline would be present for a dollarized government.

If dollarization does undermine the incentives for fiscal responsibility, does that mean it should be accompanied by legal restrictions on the government budget deficit? In part, this was the route taken by the European Monetary Union (EMU) in tying the Stability Pact to the launch of the euro. It is important to notice, however, that a unilateral dollarization is very different from the EMU's in this regard. In the latter case, members relinquished control to a common central bank for the conduct of monetary policy. The decision to dollarize, in contrast, entails total loss of monetary authority. As such, legal restrictions on the government budget would constrain an already shrunken set of policy alternatives, which could prove very costly in an economic downturn.[12]

Economic Integration

A potential benefit of dollarization is that it could increase the level of integration of the dollarizing economy with the U.S. economy. This may come about for several reasons, including reduced transactions costs and the elimination of uncertainty

[11] On the optimal-tax property of inflation, see Calvo and Guidotti (1993). For the analysis of time-consistency problems, see Kidland and Precott (1977) and Calvo (1978).

[12] Also, see Ghiglino and Shell (2000) for a discussion of when deficit restrictions do not really constrain the government and hence have no real effects.

about exchange rates. Frankel and Rose (2000) present evidence that currency unions lead to large increases in trade flows between member countries.[13] Furthermore, Frankel and Rose argue that these increases do not come from the diversion of trade away from non-member countries; rather, currency union membership leads to a higher ratio of total foreign trade to GDP. In fact, they interpret their results as indicating that increased trade is the *primary* benefit of joining a currency union (or dollarizing).

In addition to increased trade, dollarization could increase the level of financial integration between the dollarizing country and the United States. Stockman (2001) focuses on the "central bank area" that would result from dollarization. He argues that this would be the most important effect of dollarization in Mexico—the Federal Reserve System would become Mexico's central bank.

This scenario would lead to changes in monetary policy (which Stockman defines broadly to include supervisory and regulatory policies) that would affect the incentives of financial intermediaries and thereby affect the levels of investment and financial integration. This change is important because the level of financial development is strongly related to economic growth and is shown in some studies to *cause* growth[14]; thus, the potential benefits are indeed large.

Other studies, however, indicate that integration should come before dollarization. For example, Bencivenga, Huybens, and Smith (2001) show that dollarization has a different impact depending on the extent of the integration between the two economies' financial markets. They show that dollarization is beneficial when capital markets are well integrated; otherwise, dollarization may be a source of volatility and indeterminacy in the economy. Hence in their model, it is the *ex ante* level of integration of capital markets that determines the benefits of dollarization.

Bencivenga, Huybens, and Smith (2001) complement and extend the traditional optimal currency literature, where it is the integration of real markets that determines the boundaries of the optimal currency area. This theory is based on the work of Mundell (1961) and specifically addresses the issue of when two economies should use the same currency. According to the theory, the key issue in determining whether two economies fall in the same optimal currency area is whether or not there is a substantial benefit of having independent monetary policy to accommodate asymmetric shocks to the economies. An optimal currency area in general is one where: (i) asymmetric shocks are not substantial, (ii) there is high mobility of factors of production, and (iii) prices are flexible. It is important to keep in mind, however, that these considerations have not been the motivation behind the current interest in dollarization.

Using the criteria of the literature on the traditional optimal currency area, it is hard to imagine Argentina being in the same optimal currency area as the United States. Even neighboring Mexico is far from perfectly integrated with the U.S. economy. The interest on dollarization fundamentally stems from the desire to bring about financial stability. The involvement of the banking sector in the recent crises underlined the importance of this issue, which is beyond the scope of the traditional optimal-currency-area model.

The Lender of Last Resort Function

A common argument against dollarization is that it would severely limit the ability of the central bank to act as a lender of last resort when the banking sector is in distress. One of the crucial roles that banks perform is maturity transformation: taking in short-term deposits and making long-term loans. This naturally puts a bank at risk if, for whatever reason, depositors have a sudden increase in their demand for liquidity and want to withdraw their money. When there is a domestic currency that can be printed freely, the central bank always has the ability to meet this liquidity demand by lending cash to the banking sector. Banks can then repay the loans when the crisis passes. In a dollarized economy, the central bank would not have unlimited resources to lend. The fear, therefore, is that giving up the ability to print currency will make these types of crises more frequent and/or more severe.

The emerging literature has shown that this concern is likely overstated for several reasons. First, the ability of the central bank to act as a lender of last resort is equally limited under fixed exchange rates and currency boards. Nevertheless, Argentina has developed several other mechanisms to deal

[13] For a critique of their result and a review of the literature in contrast with it, see Pakko and Wall (2001).

[14] King and Levine (1993) show that financial development predicts subsequent growth, and Rajan and Zingales (1998) provide evidence of causation. See also Levine (1997). Levine and Carkovic (2001) argue that the positive effects of dollarization would be indirect, working through financial development.

with liquidity crises. These include holding excess foreign reserves (above those required to back the currency in circulation), having banks contribute to a deposit insurance fund, and contracting a type of contingent credit line with foreign banks. Velde and Veracierto (1999) calculate that, together, these mechanisms cover 40 percent of total deposits. Second, as Calvo (2001) points out, central banks in industrialized countries do not generally perform their lender-of-last-resort function by printing currency; they borrow instead. This was the case, for example, in the banking crises in Sweden and Finland in 1992. Third, as proposed by Calvo (2001) and others, a "special fund" or a credit line guarantee from an international lender of last resort could be set up to guard against a large crisis that would overwhelm domestic resources. One potential source of revenue for the fund is the increase in seignorage revenue that the United States would receive when a country dollarizes. Since the fund would likely increase the stability of dollarized-country financial markets, this could be a productive (and politically acceptable) use of the funds from the U.S. point of view.[15]

Finally, several studies have identified the domestic lender of last resort as a *cause* of both excess volatility in emerging economies' financial markets and currency crises.[16] This is largely related to the moral hazard problem that such a lender can create when the supervisory and regulatory aspects of the banking system are underdeveloped. This problem was particularly severe in East Asia and is now thought to be one of the primary causes of that crisis.[17] A related problem is that the lender of last resort might not be able to take the "right" action in times of crisis because of heavy political pressure. Ennis (2000), for example, shows how such pressure may prevent the lender of last resort from implementing the optimal policy and, instead, force the use of a suboptimal inflation tax to bail out a banking sector in distress. In this context, dollarization works as an ex ante commitment not to surrender to political pressure in the event of a liquidity crisis. Antinolfi and Keister (2000) show how dollarization can be seen as a way of committing to charge a (perhaps unpopular) "penalty rate" on discount window loans during a crisis—exactly the policy advocated by Bagehot (1873). These studies indicate that dollarization can actually be seen as fixing some of the problems created by a lender of last resort.[18] Such political-economy issues have received relatively little emphasis in the literature on dollarization, and in our opinion they deserve further research.

Existing Liability Dollarization

The set of initial conditions on which dollarization would be implemented is also crucial for understanding dollarization proposals. Our analysis would be incomplete without a discussion of the current state of an economy considering dollarizing, particularly with respect to existing liability dollarization. Liability dollarization refers to domestic borrowing denominated in or indexed to a foreign currency. Both sovereign debt and private debt in emerging-market economies are often dollarized.

Our main concern in this section is private sector dollar-denominated debt, which has been growing rapidly in emerging-market economies. This includes both direct borrowing by individual firms and borrowing by the domestic banking sector. Is widespread liability dollarization an indication that an economy should officially dollarize? The answer to this question must depend on what is causing the liability dollarization to occur. Why are firms willing to borrow in a foreign currency when this creates a balance-sheet mismatch that greatly increases their vulnerability to unexpected devaluations?

Two types of explanations have been offered in the literature. The first (see, for example, Burnside, Eichenbaum, and Rebelo, 2001) is based on (implicit or explicit) government guarantees of the liabilities, especially those of the banking system. Under a fixed exchange rate regime, the interest rate on dollar loans will be lower than the domestic interest rate, the difference reflecting the possibility of devaluation. This condition leads banks to borrow in dollars. In addition, because the government guarantee implies that it will act as a residual claimant on bank

[15] Clearly, any such contribution of seignorage would be a matter for the Congress and executive branch to decide.

[16] See, for example, Chang and Velasco (2000a), Mishkin (1999), and Fischer (1999). See also Antinolfi, Huybens, and Keister (2001), which shows how a lender of last resort having the ability to print money can allow inflationary beliefs to become self-fulfilling.

[17] See Corsetti, Pesenti, and Roubini (1999) and Mishkin (1999) on this topic.

[18] But would dollarization itself find the necessary political support to be implemented? Ennis (2000) goes on to show that this is possible if the economy has a large population of international banks (i.e., banks that operate in several countries). It is interesting to note that this is, in essence, a form of financial-market integration, which we saw above (in a different context) to be a factor that is likely to increase the probability of success with dollarization.

assets in bad states of the world (in which banks go bankrupt), banks face no ex ante incentives to purchase insurance against bad states of the world. Hence, they do not hedge (sufficiently) against foreign exchange risk. In other words, the guarantee creates a moral hazard problem that leads to a fragile banking system that is overexposed to currency risk. The reason the government would provide this guarantee is that it reduces the interest rate that domestic firms pay when financing working capital from domestic banks and, therefore, has positive effects on economic growth. This benefit the government obtains is sufficient to overcome the cost of increasing the probability of a banking crisis when the exchange rate is devalued. This is an indication that official dollarization may be warranted, as it would bring this benefit without the cost.

The second type of explanation claims that liability dollarization is a result of underdeveloped domestic financial markets (see Caballero and Krishnamurthy, 2000). The underdevelopment means that firms cannot pledge their entire return to foreign investors. As a result, assets that can be used as international collateral become essential. In such an environment, individual firms choose between borrowing in local currency (which is immune to changes in the exchange rate) and borrowing in dollars (which is cheaper). Caballero and Krishnamurthy (2000) interpret borrowing in domestic currency as purchasing insurance against exchange rate fluctuations. They go on to show how competitive markets mis-price this insurance. This problem happens because, at the firm level, there are two types of collateral—internationally accepted and domestically accepted assets. At the economy-wide level, however, only internationally accepted assets are net collateral. Because firms "overestimate" the amount of collateral that they have available, they tend to purchase less insurance than would otherwise be optimal. If this is the reason for the observed liability dollarization, it is less clear that officially dollarizing would help matters. The problem of scarce internationally accepted collateral may still arise. In this case, the benefit of official dollarization is likely to be indirect—through the development of domestic financial markets and their integration with international markets.

The Effects from the Perspective of the United States

Our discussion so far has focused on the potential costs and benefits of dollarization from the viewpoint of the economy considering dollarizing. The view of dollarization from the perspective of the United States is also important. When Ecuador and El Salvador adopted the dollar, the impact on the United States was clearly minimal. It is doubtful that the same could be said about Argentina or, especially, Mexico. Two areas where a large dollarization could have an important impact on the United States are seignorage revenue and the conduct of monetary policy.

We have discussed above how dollarization entails a transfer of seignorage revenue from the dollarizing government to the United States. We have also discussed how the dollarizing country might like to either receive a share of this money or have it set aside in a fund for lender-of-last-resort functions. The second plan might receive more support in the United States, since otherwise the United States would possibly be directly involved in trying to alleviate banking crises. This possibility introduces interesting questions about the relationship between the United States and the dollarized economies that the literature has yet to explore. To the extent that the United States perceives there to be costs to having the dollar used widely, it may be reluctant to give up the benefit of the extra revenue.

The financial integration with the United States that could follow a dollarization is commonly considered to be a major benefit of dollarizing. Arguably, financial integration can prove to be a major benefit also for the U.S. economy. In addition, however, U.S. monetary policy will have stronger effects abroad, and the United States might have to take these effects into account. As an example, suppose there is a recession in a dollarized Mexico that calls for a looser policy while events in the United States call for a tighter policy. Although the United States would have the option of ignoring events in Mexico, doing so would likely cause a significant increase in the flow of illegal immigrants into the United States. Hence the optimal policy (from a selfish point of view) would likely be looser than it would have been had Mexico kept the peso.[19] In this way, it is not only the dollarizing economy that is losing monetary independence; the United States might lose some as well.

[19] For an analysis of the potential relation between dollarization and Mexican migration to the United States see Borjas and O'N. Fisher (2001). Their results indicate that the flow of illegal immigrants is more volatile when Mexican authorities adopt a fixed exchange rate, whereas the flow of legal immigrants remains unaffected.

As a final (and highly speculative) note, we observe that, if the United States benefits from the increase in seignorage revenue, widespread dollarization would give an incentive to generate a higher steady-state level of inflation. Although it seems unlikely that this incentive would influence U.S. policy, it is interesting to report how Fischer (1982) concludes his paper:

> Use of a foreign money also implies that the domestic government is relying on the foreign government to maintain better control over the inflation rate than it does itself—an admission that most governments would be reluctant to make. And besides, Who is to guard the guardians?

FURTHER READING

We have discussed some of the key issues that are important for a country considering official dollarization, including some of the likely costs and benefits. A crucial issue that we have not discussed, however, is how large these costs and benefits would be. There is little historical evidence that can be used as guidance on this question. There are many inherent difficulties in quantifying the effects of dollarization, and these are reflected in a wide range of predictions that are obtained from different models that focus on different aspects of the problem. An example of this disparity can be found in the results of Cooley and Quadrini (2001), Del Negro and Obiols-Homs (2001), Mendoza (2001), and Schmitt-Grohé and Uribe (2001), all of which are quantitative studies related to dollarization in Mexico. Some of these papers conclude that the overall benefits would be very large, while others conclude they would be small or even negative.

All four of these papers, along with some others we have referenced and some we have not, are gathered together in a special issue of the *Journal of Money, Credit, and Banking* (May 2001). We encourage the interested reader to consult this source directly for a more extensive discussion of the issues related to dollarization than is possible here. In addition, Spanish-speaking readers are encouraged to consult *La Dolarización como Alternativa Monetaria para México* (Del Negro et al., forthcoming). This volume consists largely of papers presented at a conference on dollarization sponsored by the Instituto Technológico Autónomo de México (ITAM) in December 2000.

REFERENCES

Antinolfi, Gaetano; Huybens, Elisabeth and Keister, Todd. "Monetary Stability and Liquidity Crises: The Role of the Lender of Last Resort." *Journal of Economic Theory*, 2001, *99*(1-2), pp. 187-219.

_____ and Keister, Todd. "Liquidity Crises and Discount Window Lending: Theory and Implications for the Dollarization Debate." Working Paper 00-02, Centro de Investigación Económica, September 2000.

Bagehot, Walter. *Lombard Street*. London: William Clowes and Sons, 1873.

Bencivenga, Valerie; Huybens, Elisabeth and Smith, Bruce D. "Dollarization and the Integration of International Capital Markets: A Contribution to the Theory of Optimal Currency Areas." *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 548-89.

Borjas, George J. and O'N. Fisher, Eric. "Dollarization and the Mexican Labor Market." *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 626-47.

Burnside, Craig; Eichenbaum, Martin and Rebelo, Sergio. "Hedging and Financial Fragility in Fixed Exchange Rate Regimes." *European Economic Review*, 2001, *45*(7), pp. 1151-93.

Caballero, Ricardo J. and Krishnamurthy, Arvind. "Dollarization of Liabilities: Underinsurance and Domestic Financial Underdevelopment." Working Paper No. 7792, National Bureau of Economic Research, July 2000.

Calvo, Guillermo A. "On the Time Consistency of Optimal Policy in a Monetary Economy." *Econometrica*, 1978, *46*(6), pp. 1411-28.

_____. "Balance of Payments Crises in Emerging Markets: Large Capital Inflows and Sovereign Governments," in Paul Krugman, ed., *Currency Crises*. Chicago: University of Chicago Press, 2000.

_____. "Capital Markets and the Exchange Rate, with Special Reference to the Dollarization Debate in Latin America." *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 312-34.

_____ and Guidotti, Pablo E. "On the Flexibility of Monetary Policy: The Case of the Optimal Inflation Tax." *Review of Economic Studies*, 1993, *60*, pp. 667-87.

_____ and Reinhart, Carmen M. "When Capital Inflows Come to a Sudden Stop: Consequences and Policy Options." Unpublished manuscript, June 1999.

_____ and _____. "Fear of Floating." Unpublished manuscript, September 2000.

Chang, Roberto. "Dollarization: A Scorecard." Federal Reserve Bank of Atlanta *Economic Review*, Third Quarter 2000, *85*(3), pp. 1-11.

_____ and Velasco, Andres. "Financial Fragility and the Exchange Rate Regime." *Journal of Economic Theory*, 2000a, *92*(1), pp. 1-34.

_____ and _____. "Dollarization: Analytical Issues." Unpublished manuscript, August 2000b.

Cooley, Thomas F. and Quadrini, Vincenzo. "The Costs of Losing Monetary Independence: The Case of Mexico." *Journal of Money, Credit, and Banking*, 2001, *33*(2, Part 2), pp. 370-97.

Corsetti, Giancarlo; Pesenti, Paolo and Roubini, Nouriel. "What Caused the Asian Currency and Financial Crisis?" *Japan and the World Economy*, 1999, *11*, pp. 305-73.

De Grauwe, Paul. *International Money*. 2nd Ed. Oxford: Oxford University Press, 1996.

Del Negro, Marco; Huybens, Elisabeth and Hernández-Delgado, Alejandro, eds., *La Dolarización como Alternativa Monetaria para México*. Mexico City: Fondo de Cultura Económica, forthcoming.

_____ and Obiols-Homs, Francesc. "Has Monetary Policy Been So Bad That It Is Better to Get Rid of It?" *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 404-33.

Edwards, Sebastian. *Real Exchange Rates, Devaluation, and Adjustment*. Cambridge, MA: MIT Press, 1989.

_____. "International Capital Flows and the Emerging Markets: Amending the Rules of the Game?" Federal Reserve Bank of Boston *Conference Series 43*, June 1999, pp. 137-57.

Eichengreen, Barry. "What Problems Can Dollarization Solve?" *Journal of Policy Modeling*, forthcoming.

Ennis, Huberto M. "Banking and the Political Support for Dollarization." Working Paper 00-12, Federal Reserve Bank of Richmond, December 2000.

Fischer, Stanley. "Exchange Rate Regimes: Is the Bipolar View Correct?" Distinguished Lecture on Economics in Government, American Economic Association Meetings, January 2001. < http://www.imf.org/external/np/speeches/2001/010601a.htm > .

_____. "On the Need for an International Lender of Last Resort." *Journal of Economic Perspective*, Fall 1999, *13*, pp. 85-104.

_____. "Seignorage and the Case for a National Money." *Journal of Political Economy*, 1982, *90*(2), pp. 295-313.

Frankel, Jeffrey A. "No Single Currency Regime Is Right for All Countries at All Times." Working Paper No. 7338, National Bureau for Economic Research, September 1999.

_____ and Rose, Andrew K. "Estimating the Effects of Currency Unions on Trade and Output." Working Paper No. 7857, National Bureau for Economic Research, August 2000.

Ghiglino, Christian and Shell, Karl. "The Economic Effects of Restrictions on Government Budget Deficits." *Journal of Economic Theory*, 2000, *94*(1), pp. 106-37.

Kaminsky, Graciela L. and Reinhart, Carmen M. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." *American Economic Review*, 1999, *89*, pp. 473-500.

King, Robert G. and Levine, Ross. "Finance, Entrepreneurship, and Growth." *Journal of Monetary Economics*, 1993, *33*(3), pp. 513-42.

Kydland, Finn E. and Prescott, Edward C. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, 1977, *85*(3), pp. 473-92.

Levine, Ross. "Financial Development and Economic Growth: Views and Agenda." *Journal of Economic Literature*, 1997, *35*(2), pp. 688-726.

_____ and Carkovic, Maria. "How Much Bang for the Buck? Mexico and Dollarization." *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 339-63.

Mack, Connie. "Dollarization and Cooperation to Achieve

Sound Money." Speech given at the Federal Reserve Bank of Dallas, 6 March 2000. < http://www.dallasfed.org/htm/dallas/events/mack.html > .

Mendoza, Enrique G. "The Benefits of Dollarization When Stabilization Policy Lacks Credibility and Financial Markets Are Imperfect." *Journal of Money, Credit, and Banking*, 2001, *33*(2, Part 2), pp. 440-74.

Mishkin, Frederic S. "Lessons from the Asian Crisis." *Journal of International Money and Finance*, 1999, *18*, pp. 709-23.

Mundell, Robert. "A Theory of Optimum Currency Areas." *American Economic Review*, 1961, *51*, pp. 657-65.

Mussa, Michael; Masson, Paul; Swoboda, Alesander; Jadresic, Esteban; Mauro, Paolo; Berg, Paolo and Berg, Andy. "Exchange Rate Regimes in an Increasingly Integrated World Economy." Occasional Paper No. 193, International Monetary Fund, August 2000.

Neely, Christopher J. "An Introduction to Capital Controls." Federal Reserve Bank of St. Louis *Review*, November/December 1999, *81*(6), pp. 13-30.

Obstfeld, Maurice and Rogoff, Kenneth. "The Mirage of Fixed Exchange Rates." *Journal of Economic Perspectives*, Fall 1995, *9*(4), pp. 73-96.

Pakko, Michael R. and Wall, Howard J. "Reconsidering the Trade-Creating Effects of a Currency Union." Federal Reserve Bank of St. Louis *Review*, September/October 2001, *83*(5), pp. 37-46.

Rajan, Raghuram G. and Zingales, Luigi. "Financial Dependence and Growth." *American Economic Review*, 1998, *88*, pp. 559-86.

Schmitt-Grohé, Stephanie and Uribe, Martín. "Stabilization Policy and the Cost of Dollarization," *Journal of Money, Credit, and Banking*, 2001, *33*(2, Part 2), pp. 482-509.

Sims, Christopher A. "Fiscal Consequences for Mexico Adopting the Dollar." *Journal of Money, Credit, and Banking*, 2001, *33*(2, Part 2), pp. 597-616.

Stockman, Alan C. "Optimal Central Bank Areas, Financial Intermediation, and Mexican Dollarization." *Journal of Money, Credit, and Banking*, May 2001, *33*(2, Part 2), pp. 648-66.

Summers, Lawrence H. "International Financial Crises: Causes, Prevention, and Cures." *American Economic Review*, 2000, *90*(2), pp. 1-16.

_____. Testimony before the Senate Foreign Relations Subcommittee in International Economic Policy and Export/Trade Promotion, 27 January 1999.

Velde, François R. and Veracierto, Marcelo. "Dollarization in Argentina." *Chicago Fed Letter* No. 142, June 1999.

Persistence, Excess Volatility, and Volatility Clusters in Inflation

Michael T. Owyang

Three key features of the U.S. inflation time series have been identified by empirical studies. First, innovations in the level are persistent—that is, changes in the inflation rate generally endure. Second, volatility appears in clusters directly after changes in inflation. Following an innovation in the inflation rate, short periods of increased volatility are indicated by the presence of autoregressive conditional heteroskedasticity (ARCH) in the regression residuals. Third, periods with high mean inflation have a correspondingly high variance of inflation, and vice versa. Likewise, periods with low levels of inflation tend to be associated with low variability.

This paper presents a model in a single, integrated framework that offers one possible explanation for these facts about the U.S. inflation time series. In this model, the policymaker faces a trade-off between inflation and unemployment in the form of an expectations-augmented Phillips curve.[1] The Phillips curve is subject to two shocks: a persistent shock that follows a Markov process and a white noise shock. The magnitudes of both shocks are unobservable, forcing the policymaker to employ an ordinary least-squares (OLS) learning technology to determine the policy target. Agents have the same information set as the policymaker and form rational expectations of monetary policy. In addition to the Phillips curve shocks, an independent Markov process governs the policymaker's preferences; agents form their expectations after observing current policymaker preferences.

Learning and the interaction between the Markov processes governing the position of the Phillips curve and the policymaker's preferences provide one possible explanation for the three stylized facts about the U.S. inflation time series. Changes in the variable that determines the position of the Phillips curve (henceforth called the *structural* variable) are persistent and directly determine the policymaker's target. Thus, regime shifts in the structural variable induce persistent shifts in inflation—the first of three stylized facts. The second, volatility clustering, is driven by the policymaker's learning mechanism. Once the economic fundamentals change, the policymaker resets the learning algorithm to determine the magnitude of the new shock. As the policymaker learns, new information each period does not lead to as large of an update of his estimate of the position of the Phillips curve. Thus, periods of volatility follow the shock, then drop off.

The third fact, the relationship between mean and variance, is a result of the process governing the policymaker's preferences. In addition to persistence effects, uncertainty in the policymaker's estimate of the position of the Phillips curve produces variability in the policy target. When the policymaker is accommodative, i.e., inflation is high, this uncertainty is amplified. However, when the policymaker is in an inflation-intolerant regime, uncertainty about fluctuations in the Phillips curve are not amplified, as the policymaker is less willing to trade off inflation for small gains in unemployment.

The model is estimated using Gibbs sampling. The estimation procedure will generate both a parameter vector for the model and posterior densities for each Markov process. Monte Carlo simulations using the estimated parameters reveal that three-state versions of both Markov processes (governing the Phillips curve and the policymaker's preferences) produce artificial data that exhibit the three characteristics described above.

FEATURES OF THE U.S. DATA

Figure 1 shows the U.S. monthly annualized inflation rate for the period 1947:01–1998:05.[2] The inflation data over this period exhibit three particular time-series characteristics:

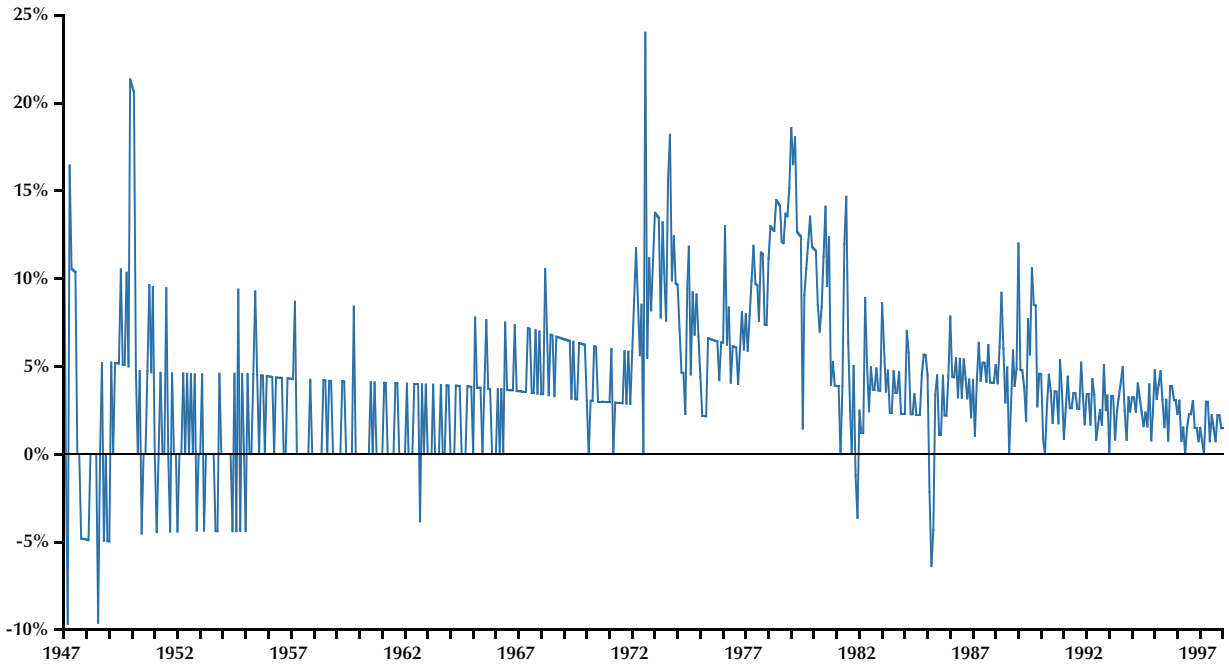
Michael T. Owyang is an economist at the Federal Reserve Bank of St. Louis. A previous version of this paper was written at the University of California, San Diego, and appears in Owyang (2000). The author thanks Jim Hamilton, Garey Ramey, and Valerie Ramey for their input, as well as Takeo Hoshi, Neal Beck, Thomas Sargent, Robert Shimer, Leora Freidberg, Kevin Lansang, Glenn Rudebusch, and seminar participants at the Federal Reserve Banks of Dallas, St. Louis, and San Francisco, the Board of Governors of the Federal Reserve System, and Emory University for useful comments. The author also thanks Chang-Jin Kim and Charles Nelson for assistance with the program. Abbigail J. Chiodo provided research assistance.

[1] This literature stems from the seminal work of Kydland and Prescott (1977) and Barro and Gordon (1983). Models possessing policymaker preferences of similar form include Barro (1986), Cukierman and Meltzer (1986), and Ball (1992, 1995).

[2] The data are seasonally adjusted CPI data taken from Citibase.

Figure 1

U.S. Inflation



1. Changes in inflation are persistent.
2. Inflation series have volatility clusters and can be modeled as some form of ARCH series.
3. High/low levels of inflation are associated with relatively high/low variance and uncertainty.

The first of these phenomena has been considered in the empirical literature by Barsky (1987) and Fuhrer and Moore (1995). They find that innovations in the rate of inflation are largely permanent, causing persistent shifts in trend; in the presence of an aggregate shock, the inflation rate rises and stays high for an extended period. Consider the subperiod 1968-73, for example; here, an innovation in the inflation rate is associated with a largely permanent shift in trend.[3]

Table 1 shows the results from an analysis of the autocorrelations for monthly data over the sample period 1947:01–1998:05. Autocorrelations for the first eight lags range between approximately 0.4 and 0.6, providing strong evidence of the presence of serial correlation. Results indicate that inflation is largely persistent with lagged coefficients that are clearly significant, indicating that innovations that occur in any period spill over into subsequent periods.

Following Engle's (1982) original analysis of the U.K. inflation data, a number of papers have attempted to fit the U.S. inflation data to an ARCH model to test for the presence of volatility clusters. Kim (1993) tests ARCH against alternate specifications and finds that ARCH does not perform as well as an unobserved-component time series with Markov-switching heteroskedasticity. However, Baillie, Chung, and Tieslau (1996) employ an alternate specification to model the inflation process, using an autoregressive, fractionally integrated, moving-average version of a generalized autoregressive conditional heteroskedasticity model (ARFIMA-GARCH). They find evidence of persistence and mean reversion, as well as heteroskedasticity, in the inflation time series.

In light of this evidence, consider a GARCH(1,1) model for the variance of the form

$$\sigma_t^2 = \kappa + \chi e_{t-1}^2 + \delta \sigma_{t-1}^2 ,$$

in which the current conditional variance depends

[3] This period coincides with the beginning of the two-tiered system for gold coverage in 1968, the closing of the gold window in 1971, and the end of the adjustable peg in 1973.

Table 1

Autocorrelations and AR(4) Regression for U.S. Time Series Inflation 1947:01–1998:05

Autocorrelations			
Lag 1	0.611	**Lag 5**	0.473
Lag 2	0.575	**Lag 6**	0.465
Lag 3	0.501	**Lag 7**	0.463
Lag 4	0.466	**Lag 8**	0.475

Variable	Coefficient	Standard error	*t* Statistic
ΔCPI_{t-1}	0.392983	0.044755	8.780743
ΔCPI_{t-2}	0.217638	0.053232	8.780741
ΔCPI_{t-3}	0.153694	0.048365	3.177771
ΔCPI_{t-4}	0.180432	0.044039	4.097130
C	0.334763	0.114108	2.933746
ARCH(1)	0.141611	0.017654	8.021489
GARCH(1)	0.830818	0.020444	40.63829

Summary statistics			
R^2	0.382133	**Mean inflation**	4.099219
Adjusted R^2	0.376006	**SD inflation**	4.511847
SE of regression	3.564057	**Akaike info criterion**	2.553171

LM test			
F Statistic	62.33962	**Probability**	0.00000
Obs R^2	56.66900	**Probability**	0.00000

Variable	Coefficient	Standard error	*t* Statistic
C	6.959891	1.016461	6.847182
RESID^2(–1)	0.302882	0.038361	7.895545

on the lagged conditional variance and the lagged squared residual. A test for ARCH by employing a Lagrange multiplier (LM) test indicates that the null hypothesis of ARCH cannot be rejected. Table 1 contains the results from the LM test of an AR(4) model and the variance results of the GARCH(1,1) regression of the U.S. inflation time series. Both the ARCH and GARCH components of the variance equations are significant.

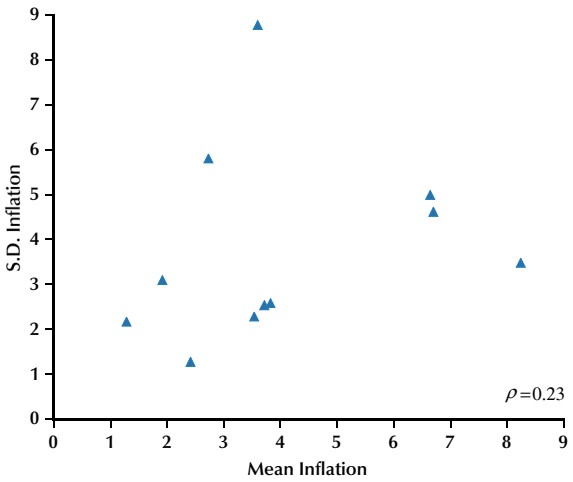
Ball and Cecchetti (1990) cite a relationship between the level of inflation and its variance, noting that an increase in the level of inflation is not only persistent but often associated with a corresponding increase in the variance and/or uncertainty of future inflation.[4] Explanations for this phenomenon have focused on three primary areas: changes in the expectations-augmented Phillips curve, temporary and permanent aggregate shocks, and idiosyncratic policy.

Using the U.S. quarterly gross national product (GNP) deflator and consumer price index (CPI) data divided into subperiods of various lengths, Ball and Cecchetti (1990) test the hypothesis that the level and variability of inflation are related. Their result is that the correlation between level and variance rises for the first few years and then begins to fall.

[4] Logue and Willet (1976), Cosimano and Jansen (1988), Devereux (1989), and Evans (1991) also test this relationship. Evans pays particular attention to estimating inflationary uncertainty. Ball (1992) provides a theoretical model that attempts to explain this correlation.

Figure 2

Correlation Between Mean and Variance for Five-Year Subperiods



Ball and Cecchetti decompose inflation into a series of permanent shocks and temporary shocks—a trend stationary component and a white noise component. They show that permanent shocks have increasing variance with level and thus cause a rise in uncertainty when trend rises.

Consider the mean and uncertainty of inflation over five-year, non-overlapping subperiods and the correlation between sample means and standard deviations across these subperiods. The results are plotted in Figure 2. Examination of the results reveals a relationship between mean and standard deviation, with a correlation of about 0.23. Figure 3 plots the inflation means and standard deviations for each subsample. The magnitude of these increases in mean and variance varies with the sample period, but this indicates a relationship that might not be completely revealed using a simple correlation test. Note that only for the period 1975-79 does the variance fall when the subsample mean rises.

Model

To formulate an integrated model that provides one possible explanation for the previously mentioned time-series characteristics of U.S. inflation data, I propose a reduced-form macroeconomic model with the following features:

- a neoclassical rational expectations–augmented Phillips curve subject to a Markov shock to the natural rate of unemployment,
- a monetary policymaker with Markov-switching preferences for low inflation relative to low unemployment, and
- policymaker learning.

The policymaker faces a short-run tradeoff between inflation and unemployment embodied in the Phillips curve.[5] Mankiw (2000) argues for the inclusion of both expectations and supply shocks to provide a complete and stable view of the economy. This model incorporates these features but assumes that the magnitude of these supply shocks is unobservable and must be learned.[6] This model focuses on the actions of a policymaker under uncertainty and subject to shifts in preferences. During each period, events occur in the following order:

- The policymaker's preferences are determined and revealed to the public.
- Agents and the policymaker simultaneously set expectations and the policy target, respectively.
- The economic shocks then occur, and the policymaker and agents observe the realized inflation and unemployment rates.
- The policymaker's and agents' information sets are updated.

Consider an economy in which a monetary policymaker sets an inflation target and private agents form expectations about the realized inflation rate.[7] Suppose that the policymaker faces an expectations-augmented Phillips curve of the form

$$u_t = k(\pi_t^e - \pi_t) + \eta_t + \varepsilon_t, \quad (1)$$

where u_t is the unemployment rate; π_t is the inflation rate; π_t^e are the period t inflation expectations; η_t is a persistent unemployment shock that follows

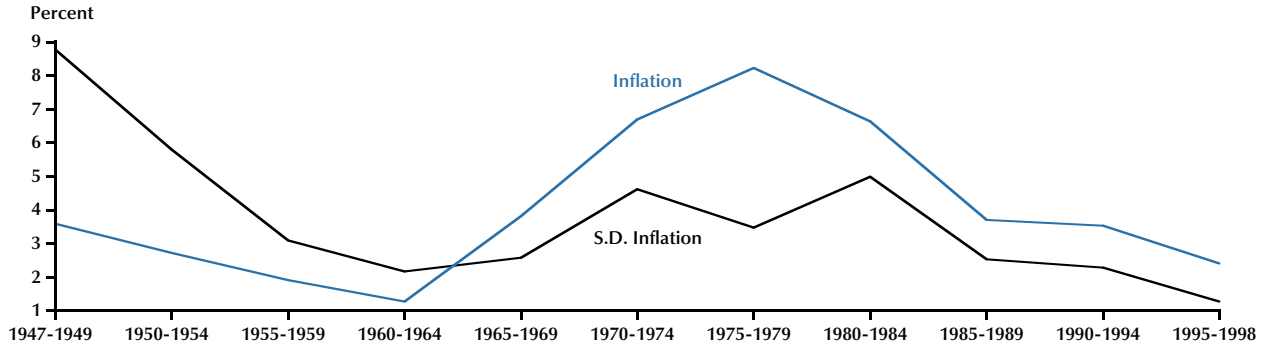
[5] In order to abstract from interest rate properties, this paper does not consider directly the policymaker's instrument. In addition, I do not include smoothing as a policymaker objective. If interest rates were added as the policymaker's instrument, the paper could be nested as a special case of the model found, for example, in Clarida, Gali, and Gertler (1999). The backward-looking nature of this model, driven by the learning process in agent expectations, is similar to that proposed originally by Taylor (1981) and later by Fuhrer and Moore (1995) and Rudebusch and Svenson (1998).

[6] A number of academic papers consider the effect of the Fed learning about the world. Sargent (1999) considers a model in which uncertainty generates paths between economic equilibria. Kasa (1999), Lansang (2001), Sack (1998), and Wieland (1998, 2000) also consider models in which the policymaker faces some degree of parameter or model uncertainty.

[7] Owyang and Ramey (2001) use a similar model with adaptive expectations to measure monetary policy.

Figure 3

Mean and Variance of Five-Year Periods Over Time



a three-state Markov-switching process with a vector of possible structural states $\vec{h} = (h_1, h_2, h_3)$, a transition kernel T_η, and period t state indicator vector S_t; and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is a white noise shock that occurs each period.

Shifts in η_t represent dramatic, persistent changes in the current economic environment. These shifts can be viewed as unexpected but highly visible events, such as wars or oil crises, in which the timing of the event is clearly observed.[8] However, the new *value* of η_t that arises following a switch cannot be observed, but rather must be learned by the policymaker and agents.[9] Let $\hat{\eta}_t = E_t[\eta_t | \Omega_t]$ denote the policymaker's period t estimate of η_t conditional on his information set $\Omega_t = (u_1, u_2, \ldots, u_t, \pi_1, \pi_2, \ldots, \pi_t)$.

The policymaker has preferences over inflation and unemployment, $L_t = \alpha_t u_t^2 + \pi_t^2$, with the relative weight on unemployment, α_t, governed by a Markov process with transition kernel T^α and possible states $\vec{\alpha} = (a_1, a_2, a_3)$. A high α_t policymaker attaches more weight to output and sets a higher inflation target, conditional on the estimated state of the world, $\hat{\eta}_t$. The low α_t policymaker is an inflation hawk and sets a lower conditional inflation target. It can be shown that a policymaker who minimizes the current period value of his quadratic loss will form a short-run, discretionary inflation target that can be given by

$$\bar{\pi}_t = k\alpha_t \frac{(k\pi_t^e + \hat{\eta}_t)}{1 + k^2\alpha_t}, \tag{2}$$

where $\hat{\eta}_t$ is the policymaker's estimate of the magnitude of the unemployment shock.[10] Although the policymaker can set an inflation target, the realized inflation rate is subject to some normally distributed noise, υ_t.[11]

For exposition, assume that agents form expectations rationally and with the same information set, Ω_t, as the policymaker. Agent expectations are then

$$\pi_t^e = \bar{\pi}_t . \tag{3}$$

Combining equations (2) and (3) gives

$$\bar{\pi}_t = k\alpha_t \hat{\eta}_t . \tag{2′}$$

Equation (2′) can be interpreted in terms of current and historical policy. Assuming that $\alpha_t \neq 0$, the policymaker's short-run target moves as a function of his belief about the state of the economy. As the balance of risks shifts toward higher perceived

[8] It is convenient but not necessary to assume that the policymaker observes the timing of the event. Similar results obtain if the event is unobserved. If there is a switch but the policymaker believes there is none, the policymaker's inflation target will induce a poor unemployment outcome. If this persists for a number of periods, the policymaker can conclude that a switch has occurred. On the other hand, if the policymaker believes a switch occurred when it in fact had not, the policymaker resets the gain sequence. New information does not change the target on average but does induce more volatility.

[9] Agents in the model, including the policymaker, are not assumed to know the structure of the underlying process. They are assumed to believe that the economy can take on a continuum of possible states.

[10] Note that this formulation assumes the policymaker is optimizing over the current period only. A more forward-looking policymaker complicates the learning via feedback from the economy and is left as an extension. This formulation, however, does not presuppose a lack of a consistent long-run inflation target.

[11] This can be interpreted as the policymaker setting an intermediate interest rate target and realizing an inflation target with some error. For this application, I suppress the interest rate target and consider only the effect on inflation.

unemployment (i.e., higher $\hat{\eta}_t$), the policymaker will allow the inflation rate to wander while focusing attention on achieving a higher growth rate (and, thus, lower unemployment) and vice versa.

The effect of the switching process α_t on the inflation target can be historically exemplified by changes in Federal Reserve chairman, although the model does not restrict it to be so. Clarida, Gali, and Gertler (2000), among others, recognize the fundamental change in Fed objectives at the onset of the Volcker regime in October 1979.

Finally, since the magnitude of the unemployment shock is unknown, the policymaker must infer the state of the economy from the data. Since the timing of the shock is known, the policymaker can employ an OLS learning technology of the form

$$\hat{\eta}_t = \frac{1}{g_t} u_{t-1} + \frac{g_t - 1}{g_t} \hat{\eta}_{t-1}, \tag{4}$$

where g_t is the number of periods since the last switch in η_t. The sequence is $g_1, g_2, \ldots, g_t$ referred to as the *gain sequence*. The policymaker resets the gain sequence when a switch occurs. This is synonymous with discarding previous (and now useless) information gathered before the shock and attaching more weight to incoming information. As the policymaker accumulates more information, new information becomes less valuable and the weight attached falls.

Inflation Dynamics

The model specified by equations (1) through (4) simplifies into two time-series equations governing unemployment and inflation: the Phillips curve (equation (1)) and

$$\pi_t = \frac{g_t - 1}{g_t} \overline{\pi}_{t-1} + \frac{k\alpha_t}{g_t} (\eta_{t-1} + \varepsilon_{t-1} - \upsilon_{t-1}) + \upsilon_t , \tag{5}$$

where υ_t is normally distributed inflation noise that occurs after the policymaker sets the inflation target and can be interpreted as control error.[12]

Equation (5) embodies the aforementioned time-series characteristics of the inflation time series. First, the autoregressive nature of equation (5) and the Markov structure of the shocks indicate that shocks to the inflation rate in the form of preference shocks to α_t or structural shocks to η_t are persistent, i.e., shifts in the inflation rate are lasting. This can be verified if the diagonal elements of the transition kernel sum to greater than 1. Second, a Markov shock to the Phillips curve in the form of a shift in η_t will induce periods of volatility while the policymaker learns. Third, an accommodative policymaker (i.e., high α_t) will tend to induce more volatility in the inflation rate than the relatively hawkish (i.e., low α_t) policymaker.[13]

The first property implies that the equilibrium inflation path exhibits persistence whenever $\alpha_t > 0$, as a direct result of persistence in the processes determining α_t and η_t and indirectly through the effect of the learning rule on the policy choices. The second property is demonstrated in the following thought experiment. Consider an aggregate shock at time $t = 0$ that shifts the Phillips curve out through an increase in the parameter η_t. The policymaker would react to the shock by raising his inflation target and raising the mean level of inflation, thus causing a persistent change in the inflation rate. The policymaker, however, does not know the exact value of the target and cannot infer the magnitude of the shock because of unemployment noise, ε_t, and inflation noise, υ_t. He must therefore implement a policy based on estimates of the new Phillips curve parameters, which he constructs using the learning technology. In the short term, the policymaker's estimate—and thus the inflation rate—is greatly affected by new information in the form of new observations. This result stems from the policymaker resetting the gain sequence, g_t, in equation (4). The weight on new information is increased because information prior to the incidence of the new shock is no longer valuable. Over time, the policymaker collects inflation-unemployment data and updates his estimate of the shock magnitude, causing the accuracy of his prediction to increase. Also, the accumulation of information allows the policymaker to decrease the weight of new information, and inflation becomes less variable.[14]

Suppose now that the policymaker's preferences change, switching from an accommodating regime to an anti-inflationary regime. The policymaker attaches more emphasis to minimizing the level of inflation and subsequently reduces the inflation

[12] Equation (5) elucidates a time-series path for inflation. A similar equation could be written for unemployment. However, the time-series characteristics for unemployment are less well documented, and I leave discussion of them to another paper. It is sufficient to say that, in this formulation, the unemployment time series would follow a Markov process with noise.

[13] A formal presentation of these results is given in Appendix A.

[14] This paper assumes a particular learning specification. However, these results are robust to permutations of the policymaker's learning technology provided that the learning mechanism implemented converges (see Marcet and Sargent, 1989).

Table 2

Estimated Parameters

Parameter	Estimated value	Parameter	Estimated value
a_3	1.2130 (0.3681)	h_3	6.1817 (1.2968)
a_2	0.4856 (0.2468)	h_2	5.3110 (0.9644)
a_1	0.0929 (0.1516)	h_1	4.3160 (0.9168)
σ_υ	2.7434 (1.1105)	σ_ε	2.4194 (1.0503)
$\Pr[\alpha_t = a_3 \mid \alpha_{t-1} = a_3]$	0.9663 (0.0191)	$\Pr[\eta_t = h_3 \mid \eta_{t-1} = h_3]$	0.9908 (0.0398)
$\Pr[\alpha_t = a_2 \mid \alpha_{t-1} = a_2]$	0.9706 (0.0189)	$\Pr[\eta_t = h_2 \mid \eta_{t-1} = h_2]$	0.7645 (0.2057)
$\Pr[\alpha_t = a_1 \mid \alpha_{t-1} = a_1]$	0.9058 (0.0768)	$\Pr[\eta_t = h_1 \mid \eta_{t-1} = h_1]$	0.6874 (0.2133)

NOTE: Standard deviations across iterations are given in parentheses.

target for a given estimate, $\hat{\eta}_t$. Additionally, because the policymaker's preferences enter multiplicatively instead of additively, the adjustments made during the learning process become smaller. Thus, the policymaker sets a lower target with less variability—the third property.

ESTIMATION

The preceding model of inflation and unemployment is estimated using Gibbs sampling. Gibbs sampling uses an iterative filtering algorithm and a Monte Carlo algorithm to generate the ergodic density for the parameter vector conditional on the data. An outline of the sampling procedure appears in Appendix B. Seasonally adjusted monthly data for both series are taken from Citibase for the sample period 1947:01–1998:05. The inflation rate is taken as the annualized rate of change of the CPI.[15] The parameter estimates are given in Table 2.

Here, $\vec{\alpha} = (a_1, a_2, a_3)$ reflects the policymaker's preferences and $\vec{h} = (h_1, h_2, h_3)$ describes unemployment shocks to the Phillips curve. The combination (a_i and $\Pr[\alpha_t = a_i \mid \alpha_{t-1} = a_j]$ for $i,j = 1,2,3$) defines the Markov process that determines the policymaker's preferences. Recall that α directly affects the policymaker's inflation target. The high value for the preference parameter, a_3, indicates the most accommodative policymaker, while the low value, a_1, represents a policymaker less willing to trade high inflation for lower unemployment; a_1 defines a regime in which the policymaker sets a nearly zero inflation target that does not respond much to changes in the state of the economy. On the other hand, a_2 and a_3 correspond to policymakers who are increasingly responsive to shocks to the structural variable.

Suppose a shock hits the economy in the form of an increase in the natural rate η_t. The accommodative policymaker ($\alpha_t = a_3$) responds to the shock by increasing his inflation target. A 1 percent increase in the natural rate, η_t, when the policymaker is in the a_3 state implies a 1.3 percent increase in the inflation target. Similarly, a 1 percent increase in the natural rate when the policymaker is in the a_1 state implies a 0.5 percent increase in the inflation target.

The parameters ($\Pr[\alpha_t = a_i \mid \alpha_{t-1} = a_i]$ for $i = 1,2,3$) are the on-diagonal transition probabilities that determine the duration of the policymaker's preference regime. They also determine the expected number of regime changes over the sample period. The number of regime shifts predicted by the estimated transition probabilities are similar to the predictions in Romer and Romer (1989).[16]

Now consider the three-state process for the structural variable η_t. The process (h_i, $\Pr[\eta_t = h_i \mid \eta_{t-1} = h_j]$ for $i,j = 1,2,3$) represents three shocks to the natural rate of unemployment and their transition probabilities. One interpretation of this process is that the state $\eta_t = h_2$ represents the "normal" state of the economy, while the other two states are the product of exogenous shocks. The low value, h_1, can be interpreted as the natural rate of unemployment in an unusually productive state of the economy,

[15] Results using the monthly personal consumption expenditures (PCE) index are not appreciably different. Postwar quarterly GDP deflator data are of insufficient length to run this type of estimation.

[16] Romer and Romer find seven instances during this sample when the Fed reacted to explicitly reduce inflation. For a more detailed comparison of the type of model presented here and the Romer dates, see Owyang and Ramey (2001).

Table 3

Three-State Monte Carlo Results

	Simulated*	U.S. time series†
AR(1)	0.3114 (0.0454)	0.3841 (0.0438)
AR(2)	0.2326 (0.0401)	0.2238 (0.0517)
AR(3)	0.1886 (0.0436)	0.1604 (0.0468)
AR(4)	0.1720 (0.0389)	0.1739 (0.0436)
Constant	3.2497 (2.883)	0.2554 (0.1053)
ARCH	0.0400 (0.0488)	0.1410 (0.0170)
GARCH	0.6753 (0.2754)	0.8406 (0.0195)
Five-year correlation‡	0.2453 (0.3660)	0.3155
Sample mean	3.89 (0.21)	3.99
Variance	4.30 (0.61)	4.40
Skewness	0.92 (0.15)	0.85
Kurtosis	2.70 (0.21)	5.02

NOTE: Simulated data statistics are taken from 1000 iterations of 608 period samples. The first column contains results from the regression of inflation on four lags. GARCH(1,1) parameters are taken from the following model of the variance $\sigma_t^2 = \kappa + \chi u_{t-1}^2 + \delta\sigma_{t-1}^2, \chi, \delta > 0$.

*Standard deviations across samples are given in parentheses.

†Standard errors for the AR(4) regression are given in parentheses.

‡Indicates the correlation between mean and variance of five-year intervals.

perhaps caused by a positive technology shock. The high value, h_3, represents the innovation to the natural rate after an adverse event such as an oil shock.[17]

Simulations

To evaluate how well the proposed model fits the data, the estimated parameters from the previous section are used in Monte Carlo simulations. The goal is to determine whether the model replicates the moment and variability characteristics of the U.S. inflation time series. Artificial data series are created according to equation (5), and AR(4) regressions are performed for comparison with the U.S. time series.[18] The residuals are then tested for GARCH(1,1). Results of the simulations and subsequent statistical tests are shown in Table 3.

The coefficients on the autoregressive terms and variability characteristics taken from the regressions on the artificial data are statistically consistent with those obtained from the U.S. time series. However, both the ARCH and GARCH parameters are comparatively low and more variable. This can be explained by the manner in which the simulations are generated. The estimated variability characteristics show dependency on the average duration of the regime. Short-lived regimes in η_t will reduce the variability characteristics of the artificial data, as the policymaker is unable to carry out the learning process. The policymaker resets the gain at the onset of each switch; however, if regimes tend to be of short duration, the policymaker never has a chance to reduce the gain. Variability characteristics are driven by the switches in the regime rather than the learning process. Long regimes in the η_t variable can cause similar problems. When there are few switches, the policymaker's gain remains at low levels, reducing the amount of variability induced by learning. Analysis of individual simulations reveals that, in iterations in which the policymaker reverts to an anti-inflationary regime for unusually long periods, the GARCH coefficient is near zero. Essentially, the variability parameters are biased downward in both cases of extremely high numbers of switches and extremely low numbers of switches. Incorporating switching processes into Monte Carlo experiments of this kind makes these results unavoidable, as some samples are bound to be outliers.

The artificial data also exhibit the mean-variability relationship found in the U.S. time series. The correlation between mean and variance for five-year intervals is highly variable in the simulated data, however. This correlation is also caused by the incorporation of Markov processes into Monte Carlo experiments. Recall that the mean-variability relationship depends on switches in the preference process. Thus, the statistic will be biased downward if the changes in the inflation level are driven primarily by switches in the structural process. If no switches in the α_t process occur, the mean-variance correlation can actually be negative. A similar result is true if there is a large number of preference switches, which can also have a tendency to understate the relationship or cause a negative correlation.

[17] An alternate explanation for shifts in the natural rate involves the change in demographics over time. Shimer (1999) attributes recent changes in the natural rate to the "baby-boomers" phenomenon.

[18] Series lengths are 750 observations. The first 142 observations are discarded to avoid issues associated with initial conditions.

CONCLUSION

The above model combines a reduced-form model of monetary policy with a Marcet-Sargent (1989) policymaker OLS learning formulation to provide one possible explanation for the three stylized facts about U.S. inflation. Results indicate that a bivariate, three-state Markov-switching model can generate these characteristics of the U.S. inflation time series. The model is able to produce inflation persistence, volatility clusters, and a correlation between level and variance and the parameter estimates that are similar to those of the actual U.S. inflation time series.

These inflation characteristics are generated by the interaction among the unobserved shock to the Phillips curve, policymaker learning, and the switches in the policymaker's preferences. Persistent shocks to either preferences or the Phillips curve translate, through the policymaker's decision rule, into persistent changes in the inflation target. Additionally, more inflation-tolerant policymakers tend to allow shocks to the economy to have a greater effect on the inflation target, causing the inflation target to become more variable when it is relatively high. Finally, clusters of volatility are a reflection of the policymaker's learning process. The uncertainty generated by the onset of a new shock makes the target more variable; however, as the policymaker learns, the target settles down.

What might this model, if correct, imply for the future of the U.S. economy? Has the economy entered a new technological regime? Can the economy sustain a low level of unemployment? Is a slowdown inevitable? Forecasters move rapidly to change their opinions at the onset of new information. Fluctuations in economic indicators can quickly alter the tenor of expectations. The Fed scrambles to adjust rates; economists on Wall Street adjust expectations. The result: more volatility while we all figure out the real state of the economy.

REFERENCES

Baillie, Richard; Chung, Ching-Fan and Tieslau, Margie. "Analysing Inflation by the Fractionally Integrated ARFIMA-GARCH Model." *Journal of Applied Econometrics*, January/February 1996, *11*(1), pp. 23-40.

Ball, Laurence. "Why Does High Inflation Raise Inflation Uncertainty?" *Journal of Monetary Economics*, June 1992, *29*(3), pp. 371-88.

_____. "Time-Consistent Policy and Persistent Changes in Inflation." *Journal of Monetary Economics*, November 1995, *36*(2), pp. 329-50.

_____ and Cecchetti, Stephen. "Inflation and Uncertainty at Short and Long Horizons." *Brookings Papers on Economic Activity*, 1990, (1), pp. 215-54.

Barro, Robert. "Reputation in a Model of Monetary Policy with Incomplete Information." *Journal of Monetary Economics*, January 1986, *17*(1), pp. 3-20.

_____ and Gordon, David. "Rules, Discretion and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics*, July 1983, *12*(1), pp. 101-21.

Barsky, Robert. "The Fischer Hypothesis and the Forecastability and Persistence of Inflation." *Journal of Monetary Economics*, January 1987, *19*(1), pp. 3-24.

Carter, Chris K. and Kohn, Robert. "On Gibbs Sampling for State Space Models." *Biometrika*, August 1994, *81*(3), pp. 541-53.

Casella, George and George, Edward. "Explaining the Gibbs Sampler." *The American Statistician*, August 1992, *46*(3), pp. 167-74.

Clarida, Richard; Galí, Jordi and Gertler, Mark. "The Science of Monetary Policy: A New Keynesian." *The Journal of Economic Literature*, December 1999, *37*(4), pp. 1661-707.

_____; _____ and _____. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics*, February 2000, *115*(1), pp. 147-80.

Cosimano, Thomas and Jansen, Dennis. "Estimates of the Variance of U.S. Inflation Based Upon the ARCH Model: A Comment." *Journal of Money, Credit, and Banking*, August 1988, *20*(3), pp. 409-21.

Cukierman, Alex and Meltzer, Allan. "A Theory of Ambiguity, Credibility, and Inflation Under Discretion and Asymmetric Information." *Econometrica*, September 1986, *54*(5), pp. 1099-128.

Devereux, Michael. "A Positive Theory of Inflation and Inflation Variance." *Economic Inquiry*, January 1989, *27*(1), pp. 105-16.

Engle, Robert. "Autoregressive Conditional Heteroskedasticity

with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, July 1982, *51*(4), pp. 987-1008.

_____. "Estimates of the Variance of U.S. Inflation Based Upon the ARCH Model." *Journal of Money, Credit, and Banking*, August 1983, *15*(3), pp. 286-301.

Evans, Martin. "Discovering the Link Between Inflation Rates and Inflation Uncertainty." *Journal of Money, Credit, and Banking*, May 1991, *23*(2), pp. 169-84.

Fuhrer, Jeffery and Moore, George. "Inflation Persistence." *Quarterly Journal of Economics*, February 1995, *110*(1), pp. 127-59.

Hamilton, James. "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics*, July/August 1990, *45*(1-2), pp. 39-70.

Kasa, Kenneth. "Will the Fed Ever Learn?" *Journal of Macroeconomics*, Spring 1999, *21*(2), pp. 279-92.

Kim, Chang-Jin. "Unobserved-Component Time Series Models with Markov-Switching Heteroscedasticity: Changes in Regime and the Link Between Inflation Rates and Inflation Uncertainty." *Journal of Business and Economic Statistics*, July 1993, *11*(3), pp. 341-49.

_____ and Nelson, Charles. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: MIT Press, 1999.

Kydland, Finn E. and Prescott, Edward C. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, June 1977, *85*(3), pp. 473-91.

Lansang, Kevin. "Learning About a Shift in Trend Output: Implications for Monetary Policy and Inflation." Working Paper No. 00-16, Federal Reserve Bank of San Francisco, 2001.

Logue, Dennis and Willet, Thomas. "A Note on the Relation Between the Rate and Variability of Inflation." *Economica*, May 1976, *43*(17), pp. 151-58.

Marcet, Albert and Sargent, Thomas. "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models." *Journal of Economic Theory*, August 1989, *48*(2), pp. 337-68.

Mankiw, N. Gregory. "The Inexorable and Mysterious Tradeoff Between Inflation and Unemployment." Working Paper No. W7884, National Bureau of Economic Research, September 2000.

Owyang, Michael. "Three Essays on the Behavior of Monetary Policy." Ph.D. Dissertation, University of California, San Diego, 2000.

_____ and Ramey, Garey. "Regime Switching and Monetary Policy Measurement." Working Paper No. 2001-03, University of California, San Diego, 2001.

Romer, Christina D. and Romer, David H. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," in *National Bureau of Economic Research Macroeconomics Annual: 1989*. Cambridge, MA: MIT Press, 1989, pp. 121-70.

Rudebusch, Glenn D. and Svenson, Lars E.O. "Policy Rules for Inflation Targeting." Working Paper No. W6512, National Bureau of Economic Research, April 1998.

Sack, Brian. "Uncertainty, Learning and Gradual Monetary Policy." Working Paper No. 1998-34, Federal Reserve Board of Governors, August 1998.

Sargent, Thomas. *The Conquest of American Inflation*. Princeton: Princeton University Press, 1999.

Shimer, Robert. "Why Is the U.S. Unemployment Rate So Much Lower?" in *National Bureau of Economic Research Macroeconomics Annual: 1998*. Cambridge, MA: MIT Press, 1999, pp. 11-61.

Taylor, John. "On the Relation Between the Variability of Inflation and the Average Inflation Rate." Carnegie-Rochester Conference Series on Public Policy, Autumn 1981, *15*(0), pp. 57-85.

_____. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy*, February 1980, *88*(1), pp. 1-23.

Wieland, Volker. "Monetary Policy and Uncertainty About the Natural Unemployment Rate." Finance and Economics Discussion Paper Series, No. 98/22, Board of Governors of the Federal Reserve System, April 1998.

_____. "Learning by Doing and the Value of Optimal Experimentation." *Journal of Economic Dynamics and Control*, April 2000, *24*(4), pp. 501-34.

Appendix A

PROPOSITIONS AND PROOFS

Proposition 1. A shift in either α_t or η_t results in a persistent change in the inflation target $\bar{\pi}_t$.

Proposition 2. A one-time switch in the structural parameter causes the inflation variance to rise in periods following the switch.

Proof. Given equation (5) and the fact that the shocks are uncorrelated, write the variance of inflation, conditional on g_t, as

$$\text{(A1)}\quad Var(\pi_t \mid g_t) = \left(\frac{g_t - 1}{g_t}\right)^2 var(\bar{\pi}_{t-1} \mid g_t) + \left(\frac{k\alpha}{g_t}\right)^2 (\sigma_\upsilon^2 + \sigma_\varepsilon^2) + \sigma_\upsilon^2 .$$

Further

$$Var(\hat{\eta}_t \mid g_t) = \frac{\sigma_\varepsilon^2}{g_t - 1} .$$

Combining yields

$$\text{(A2)}\quad Var(\pi_t \mid g_t) = \frac{(k\alpha_t)^2}{g_t}\sigma_\varepsilon^2 + \left(\frac{k\alpha_t}{g_t}\right)^2 \sigma_\upsilon^2 + \sigma_\upsilon^2 .$$

For $t < t'$, the economy is in a steady state with $g_t = \infty$, and $Var(\pi_t|g_t) = \sigma_\upsilon^2$. Further, (A2) implies that $Var(\pi_t|g_t)$ is strictly decreasing in g_t, from which follows that the variance of inflation rises following t'. **Q.E.D.**

Proposition 3. The mean level of inflation is increasing in α. Additionally, when α is larger, the switch in η_t causes the variance of inflation to rise by a greater amount.

Proof. The first statement follows directly from (5). When α rises, the expected value of the second term on the right-hand side of (5) rises and the expected value of the next period target rises. In period t, the expected value of the second term is again larger than it was before the regime shift. The first term has also risen because the target rose that period. The conclusion follows by induction. Consider again (A1). It is easy to verify that the variance of inflation rises with α, which implies the latter result. **Q.E.D.**

Appendix B

THE GIBBS SAMPLER

The goal of the sampling routine is to estimate the conditional posterior distribution $p(\tilde{\beta}_T, \tilde{S}_T, \tilde{Z}_T, \omega|\tilde{y}_T)$, where $\tilde{y}_T = (y_1, y_2, \ldots, y_T)$ is the vector of observables through time T and ω is the vector of parameters governing both Markov processes and the variances of the white noise shocks to the Phillips curve and the inflation rate.[19] The Markov processes make the direct estimation of this distribution impractical. However, draws from $p(\tilde{\beta}_T, \tilde{S}_T, \tilde{Z}_T, \omega|\tilde{y}_T)$ can be made from an ergodic distribution of Markov simulations generated iteratively from the following conditional densities:

$$\text{(B1)}\quad p(\tilde{\beta}_T, \tilde{S}_T, \tilde{Z}_T, \omega \mid \tilde{y}_T) \begin{cases} p(\tilde{\beta}_T \mid \tilde{y}_T, \tilde{S}_T, \tilde{Z}_T, \omega) \\ p(\tilde{S}_T \mid \tilde{y}_T, \tilde{\beta}_T, \tilde{Z}_T, \omega) \\ p(\tilde{Z}_T \mid \tilde{y}_T, \tilde{\beta}_T, \tilde{S}_T, \omega) \\ p(\omega \mid \tilde{y}_T, \tilde{\beta}_T, \tilde{S}_T, \tilde{Z}_T) \end{cases} .$$

The process of drawing from each of the above marginal distributions is accomplished by a version of Carter and Kohn's (1994) multi-move algorithm, with one exception. The conditional distribution $p(\tilde{S}_T|\tilde{y}_T, \tilde{\beta}_T, \tilde{Z}_T, \omega)$ differs from the others in (B1) in that S_t depends on $\tilde{S}_{t-1}$ because of the policymaker's gain sequence. The multi-move algorithm generates the entire vector $\tilde{S}_T$ simultaneously. However, given S_t's dependence on previous states, it cannot be generated in this manner. Thus, $\tilde{S}_T$ will be drawn from sequential sampling from

$$\text{(B2)}\qquad p(S_t \mid \tilde{S}_{-t}, \tilde{y}_T, \tilde{\beta}_T, \tilde{Z}_T, \omega) ,$$

where $\tilde{S}_{-t} = (S_1, S_2, \ldots, S_{t-1}, S_{t+1}, \ldots, S_T)$. The joint density $p(\tilde{\beta}_T, \tilde{S}_T, \tilde{Z}_T, \omega|\tilde{y}_T)$ can then be estimated using the marginal densities (B1) and (B2).

[19] More detailed analyses of the Gibbs sampler are available in Cassela and George (1992) and Kim and Nelson (1999).