Economic Quarterly—Volume 93, Number 3—Summer 2007—Pages 201–227

Inflation and Unemployment: A Layperson's Guide to the Phillips Curve

Jeffrey M. Lacker and John A. Weinberg

What do you remember from the economics class you took in college? Even if you didn't take economics, what basic ideas do you think are important for understanding the way markets work? In either case, one thing you might come up with is that when the demand for a good rises—when more and more people want more and more of that good—its price will tend to increase. This basic piece of economic logic helps us understand the phenomena we observe in many specific markets—from the tendency of gasoline prices to rise as the summer sets in and people hit the road on their family vacations, to the tendency for last year's styles to fall in price as consumers turn to the new fashions.

This notion paints a picture of the price of a good moving together in the same direction with its quantity—when people are buying more, its price is rising. Of course supply matters, too, and thinking about variations in supply—goods becoming more or less plentiful or more or less costly to produce—complicates the picture. But in many cases such as the examples above, we might expect movements up and down in demand to happen more frequently than movements in supply. Certainly for goods produced by a stable industry in an environment of little technological change, we would expect that many movements in price and quantity are driven by movements in demand, which would cause price and quantity to move up and down together. Common sense

This article first appeared in the Bank's *2006 Annual Report*. The authors are Jeffrey M. Lacker, President of the Federal Reserve Bank of Richmond, and John A. Weinberg, a Senior Vice President and Director of Research. Andreas Hornstein, Thomas Lubik, John Walter, and Alex Wolman contributed valuable comments to this article. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve System.

suggests that this logic would carry over to how one thinks about not only the price of one good but also the prices of all goods. Should an average measure of all prices in the economy—the consumer price index, for example—be expected to move up when our total measures of goods produced and consumed rise? And should faster growth in these quantities—as measured, say, by gross domestic product—be accompanied by faster increases in prices? That is, should inflation move up and down with real economic growth?

The simple intuition behind this series of questions is seriously incomplete as a description of the behavior of prices and quantities at the macroeconomic level. But it does form the basis for an idea at the heart of much macroeconomic policy analysis for at least a half century. This idea is called the "Phillips curve," and it embodies a hypothesis about the relationship between inflation and real economic variables. It is usually stated not in terms of the positive relationship between inflation and growth but in terms of a negative relationship between inflation and unemployment. Since faster growth often means more intensive utilization of an economy's resources, faster growth will be expected to come with falling unemployment. Hence, faster inflation is associated with lower unemployment. In this form, the Phillips curve looks like the expression of a tradeoff between two bad economic outcomes—reducing inflation requires accepting higher unemployment.

The first important observation about this relationship is that the simple intuition described at the beginning of this essay is not immediately applicable at the level of the economy-wide price level. That intuition is built on the workings of supply and demand in setting the quantity and price of a specific good. The price of that specific good is best understood as a *relative* price—the price of that good compared to the prices of other goods. By contrast, inflation is the rate of change of the general level of all prices. Recognizing this distinction does not mean that rising demand for all goods—that is, rising aggregate demand—would not make all prices rise. Rather, the important implication of this distinction is that it focuses attention on what, besides people's underlying desire for more goods and services, might drive a general increase in all prices. The other key factor is the supply of money in the economy.

Economic decisions of producers and consumers are driven by relative prices: a rising price of bagels relative to doughnuts might prompt a baker to shift production away from doughnuts and toward bagels. If we could imagine a situation in which all prices of *all* outputs and inputs in the economy, including wages, rise at exactly the same rate, what effect on economic decisions would we expect? A reasonable answer is "none." Nothing will have become more expensive relative to other goods, and labor income will have risen as much as prices, leaving people no poorer or richer.

The thought experiment involving all prices and wages rising in equal proportions demonstrates the principle of *monetary neutrality*. The term refers to the fact that the hypothetical increase in prices and wages could be expected to result from a corresponding increase in the supply of money. Monetary neutrality is a natural starting point for thinking about the relationship between inflation and real economic variables. If money is neutral, then an increase in the supply of money translates directly into inflation and has no necessary relationship with changes in real output, output growth, or unemployment. That is, when money is neutral, the simple supply-and-demand intuition about output growth and inflation does not apply to inflation associated with the growth of the money supply.

The logic of monetary neutrality is indisputable, but is it relevant? The logic arises from thinking about hypothetical "frictionless" economies in which all market participants at all times have all the information they need to price the goods they sell and to choose among the available goods, and in which sellers can easily change the price they charge. Against this hypothetical benchmark, actual economies are likely to appear imperfect to the naked eye. And under the microscope of econometric evidence, a positive correlation between inflation and real growth does tend to show up. The task of modern macroeconomics has been to understand these empirical relationships. What are the "frictions" that impede monetary neutrality? Since monetary policy is a key determinant of inflation, another important question is how the conduct of policy affects the observed relationships. And finally, what does our understanding of these relationships imply about the proper conduct of policy?

The Phillips curve, viewed as a way of capturing how money might not be neutral, has always been a central part of the way economists have thought about macroeconomics and monetary policy. It also forms the basis, perhaps implicitly, of popular understanding of the basic problem of economic policy; namely, we want the economy to grow and unemployment to be low, but if growth is too robust, inflation becomes a risk. Over time, many debates about economic policy have boiled down to alternative understandings of what the Phillips curve is and what it means. Even today, views that economists express on the effects of macroeconomic policy in general and monetary policy in particular often derive from what they think about the nature, the shape, and the stability of the Phillips curve.

This essay seeks to trace the evolution of our understanding of the Phillips curve, from before its inception to contemporary debates about economic policy. The history presented in the pages that follow is by no means exhaustive. Important parts of economists' understanding of this relationship that we neglect include discussions of how the observed Phillips curve's statistical relationship could emerge even under monetary neutrality.[1] We also neglect the literature on the possibility of real economic costs of inflation that arise

[1] King and Plosser (1984).

even when money is neutral.[2] Instead, we seek to provide the broad outlines of the intellectual development that has led to the role of the Phillips curve in modern macroeconomics, emphasizing the interplay of economic theory and empirical evidence.

After reviewing the history, we will turn to the current debate about the Phillips curve and how it translates into differing views about monetary policy. People commonly talk about a central bank seeking to engineer a slowing of the economy to bring about lower inflation. They think of the Phillips curve as describing how much slowing is required to achieve a given reduction in inflation. We believe that this reading of the Phillips curve as a lever that a policymaker might manipulate mechanically can be misleading. By itself, the Phillips curve is a statistical relationship that has arisen from the complex interaction of policy decisions and the actions of private participants in the economy. Importantly, choices made by policymakers play a large role in determining the nature of the statistical Phillips curve. Understanding *that* relationship—between policymaking and the Phillips curve—is a key ingredient to sound policy decisions. We return to this theme after our historical overview.

1. SOME HISTORY

The Phillips curve is named for New Zealand-born economist A.W. Phillips, who published a paper in 1958 showing an inverse relationship between (wage) inflation and unemployment in nearly 100 years of data from the United Kingdom.[3] Since this is the work from which the curve acquired its name, one might assume that the economics profession's prior consensus on the matter embodied the presumption that money is neutral. But this in fact is not the case. The idea of monetary neutrality has long coexisted with the notion that periods of rising money growth and inflation might be accompanied by increases in output and declines in unemployment. Robert Lucas (1996), in his Nobel lecture on the subject of monetary neutrality, finds both ideas expressed in the work of David Hume in 1752! Thomas Humphrey (1991) traces the notion of a Phillips curve tradeoff throughout the writings of the classical economists in the 18th and 19th centuries. Even Irving Fisher, whose statement of the quantity theory of money embodied a full articulation of the consequences of neutrality, recognized the possible real effects of money and inflation over the course of a business cycle.

In early writings, these two opposing ideas—that money is neutral and that it is associated with rising real growth—were typically reconciled by the distinction between periods of time ambiguously referred to as "short

[2] Cooley and Hansen (1989), for instance.

[3] Phillips (1958).

run" and "long run." The logic of monetary neutrality is essentially long-run logic. The type of thought experiment the classical writers had in mind was a one-time increase in the quantity of money circulating in an economy. Their logic implied that, ultimately, this would merely amount to a change in units of measurement. Given enough time for the extra money to spread itself throughout the economy, all prices would rise proportionately. So while the number of units of money needed to compensate a day's labor might be higher, the amount of food, shelter, and clothing that a day's pay could purchase would be exactly the same as before the increase in money and prices.

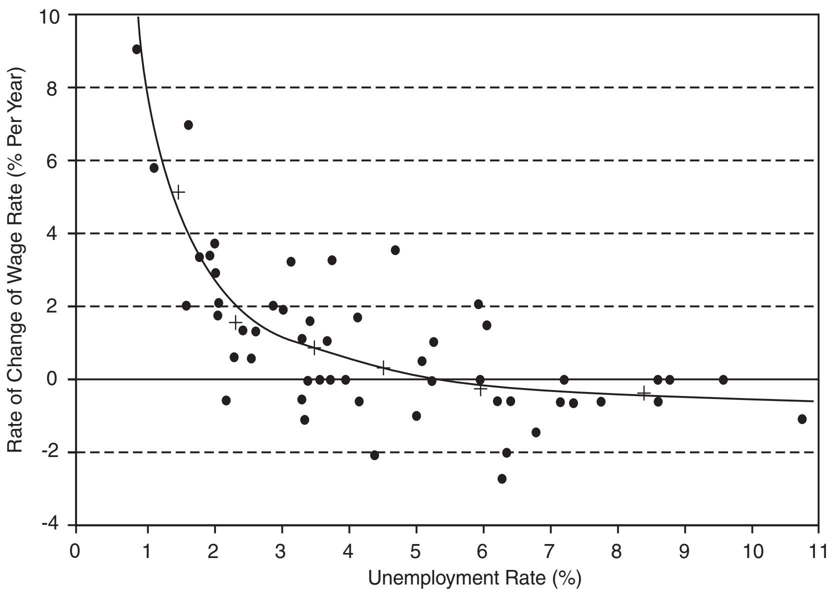
Against this logic stood the classical economists' observations of the world around them in which increases in money and prices appeared to bring increases in industrial and commercial activity. This empirical observation did not employ the kind of formal statistics as that used by modern economists but simply the practice of keen observation. They would typically explain the difference between their theory's predictions (neutrality) and their observations by appealing to what economists today would call "frictions" in the marketplace. Of particular importance in this instance are frictions that get in the way of price adjustment or make it hard for buyers and sellers of goods and services to know when the general level of all prices is rising. If a craftsman sees that he can sell his wares for an increased price but doesn't realize that all prices are rising proportionately, he might think that his goods are rising in value relative to other goods. He might then take action to increase his output so as to benefit from the perceived rise in the worth of his labors.

This example shows how frictions in price adjustment can break the logic of money neutrality. But such a departure is likely to be only temporary. You can't fool everybody forever, and eventually people learn about the general inflation caused by an increase in money. The real effects of inflation should then die out. It was in fact in the context of this distinction between long-run neutrality and the short-run tradeoff between inflation and real growth that John Maynard Keynes made his oft-quoted quip that "in the long run we are all dead."[4]

Phillips' work was among the first formal statistical analyses of the relationship between inflation and real economic activity. The data on the rate of wage increase and the rate of unemployment for Phillips' baseline period of 1861–1913 are reproduced in Figure 1. These data show a clear negative relationship—greater inflation tends to coincide with lower unemployment. To highlight that relationship, Phillips fit the curve in Figure 1 to the data. He then examined a number of episodes, both within the baseline period and in other periods up through 1957. The general tendency of a negative relationship persists throughout.

[4] Keynes (1923).

Figure 1 Inflation-Unemployment Relationship in the United Kingdom, 1861–1913



Source: Phillips (1958).

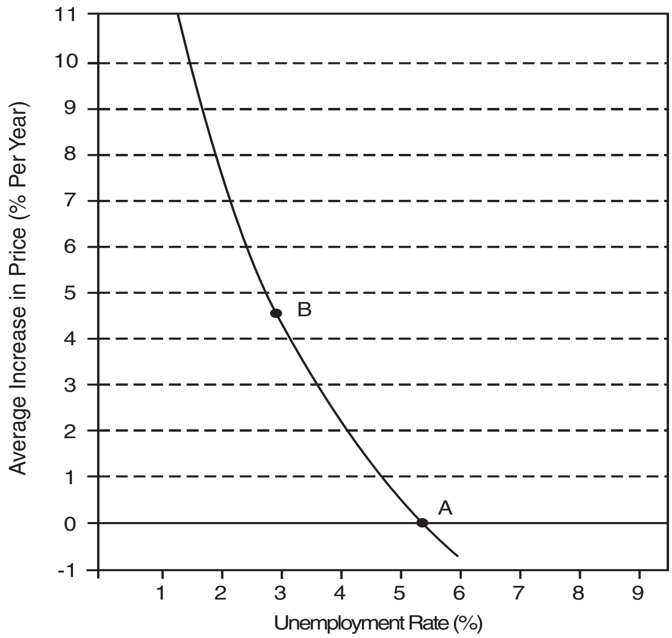
Crossing the Atlantic

A few years later, Paul Samuelson and Robert Solow, both eventual Nobel Prize winners, took a look at the U.S. data from the beginning of the 20th century through 1958.[5] A similar scatter-plot to that in Figure 1 was less definitive in showing the negative relationship between wage inflation and unemployment. The authors were able to recover a pattern similar to Phillips' by taking out the years of the World Wars and the Great Depression. They also translated their findings into a relationship between unemployment and *price* inflation. It is this relationship that economists now most commonly think of as the "Phillips curve."

Samuelson and Solow's Phillips curve is reproduced in Figure 2. They interpret this curve as showing the combinations of unemployment and inflation available to society. The implication is that policymakers must choose from the menu traced out by the curve. An inflation rate of zero, or *price stability*,

[5] Samuelson and Solow (1960).

Figure 2 Inflation-Unemployment Relationship in the United States around 1960



Source: Samuelson and Solow (1960).

appears to require an unemployment rate of about $5\frac{1}{2}$ percent. To achieve unemployment of about 3 percent, which the authors viewed as approximately full employment, the curve suggests that inflation would need to be close to 5 percent.

Samuelson and Solow did not propose that their estimated curve described a permanent relationship that would never change. Rather, they presented it as a description of the array of possibilities facing the economy in "the years just ahead."[6] While recognizing that the relationship might change beyond this near horizon, they remained largely agnostic on how and why it might change. As a final note, however, they suggest institutional reforms that *might* produce a more favorable tradeoff (shifting the curve in Figure 2 down and to the left). These involve measures to limit the ability of businesses and unions to exercise monopoly control over prices and wages, or even direct

[6] Ibid., p. 193.

wage and price controls. Their closing discussion suggests that they, like many economists at the time, viewed both inflation and the frictions that kept money and inflation from being neutral as at least partly structural—hard-wired into the institutions of modern, corporate capitalism. Indeed, they concluded their paper with speculation about institutional reforms that could move the Phillips curve down and to the left. This was an interpretation that was compatible with the idea of a more permanent tradeoff that derived from the structure of the economy and that could be exploited by policymakers seeking to engineer lasting changes in economic performance.

By the 1960s, then, the Phillips curve tradeoff had become an essential part of the Keynesian approach to macroeconomics that dominated the field in the decades following the Second World War. Guided by this relationship, economists argued that the government could use fiscal policy—government spending or tax cuts—to stimulate the economy toward full employment with a fair amount of certainty about what the cost would be in terms of increased inflation. Alternatively, such a stimulative effect could be achieved by monetary policy. In either case, policymaking would be a conceptually simple matter of cost-benefit analysis, although its implementation was by no means simple. And since the costs of a small amount of inflation to society were thought to be low, it seemed worthwhile to achieve a lower unemployment rate at the cost of tolerating only a little more inflation.

Turning the Focus to Expectations

This approach to economic policy implicitly either denied the long-run neutrality of money or thought it irrelevant. A distinct minority view within the profession, however, continued to emphasize limitations on the ability of rising inflation to bring down unemployment in a sustained way. The leading proponent of this view was Milton Friedman, whose Nobel Prize award would cite his Phillips curve work. In his presidential address to the American Economics Association, Friedman began his discussion of monetary policy by stipulating what monetary policy cannot do. Chief among these was that it could not "peg the rate of unemployment for more than very limited periods." [7] Attempts to use expansionary monetary policy to keep unemployment persistently below what he referred to as its "natural rate" would inevitably come at the cost of successively higher inflation. Key to his argument was the distinction between anticipated and unanticipated inflation. The short-run tradeoff between inflation and unemployment depended on the inflation expectations of the public. If people generally expected price stability (zero inflation), then monetary policy that brought about inflation of 3 percent would stimulate the

[7] Friedman (1968), p. 5.

economy, raising output growth and reducing unemployment. But suppose the economy had been experiencing higher inflation, of say 5 percent, for some time, and that people had come to expect that rate of increase to continue. Then, a policy that brought about 3 percent inflation would actually slow the economy, making unemployment tend to rise.

By emphasizing the public's inflation expectations, Friedman's analysis drew a link that was largely absent in earlier Phillips curve analyses. Specifically, his argument was that not only is monetary policy primarily responsible for determining the rate of inflation that will prevail, but it also ultimately determines the location of the entire Phillips curve. He argued that the economy would be at the natural rate of unemployment in the absence of unanticipated inflation. That is, the ability of a small increase in inflation to stimulate economic output and employment relied on the element of surprise. Both the inflation that people had come to expect and the ability to create a surprise were then consequences of monetary policy decisions.

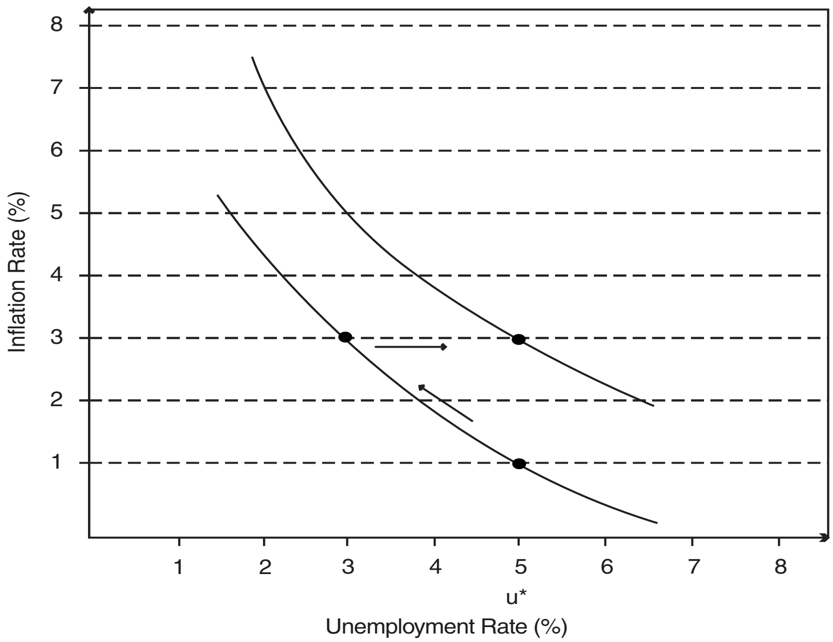
Friedman's argument involved the idea of a "natural rate" of unemployment. This natural rate was something that was determined by the structure of the economy, its rate of growth, and other real factors independent of monetary policy and the rate of inflation. While this natural rate might change over time, at any point in time, unemployment below the natural rate could only be achieved by policies that created inflation in excess of that anticipated by the public. But if inflation remained at the elevated level, people would come to expect higher inflation, and its stimulative effect would be lost. Unemployment would move back toward its natural rate. That is, the Phillips curve would shift up and to its right, as shown in Figure 3.

The figure shows a hypothetical example in which the natural rate of unemployment is 5 percent and people initially expect inflation of 1 percent. A surprise inflation of 3 percent drives unemployment down to 3 percent. But sustained inflation at the higher rate ultimately changes expectations, and the Phillips curve shifts back so that the natural rate of unemployment is achieved but now at 3 percent inflation. This analysis, which takes account of inflation expectations, is referred to as the *expectations-augmented Phillips curve*. An independent and contemporaneous development of this approach to the Phillips curve was given by Edmund Phelps, winner of the 2006 Nobel Prize in economics.[8] Phelps developed his version of the Phillips curve by working through the implications of frictions in the setting of wages and prices, which anticipated much of the work that followed.

The reasoning of Friedman and Phelps implied that attempts to exploit systematically the Phillips curve to bring about lower unemployment would succeed only temporarily at best. To have an effect on real activity, monetary

[8] Phelps (1967).

Figure 3 Expectations-Augmented Phillips Curve



Notes: When expected inflation is 1 percent, an unanticipated increase in inflation will initially bring unemployment down. But expectations will eventually adjust, bringing unemployment back to its natural rate (u^*) at the higher rate of inflation.

policy needed to bring about inflation in excess of people's expectations. But eventually, people would come to expect higher inflation, and the policy would lose its stimulative effect. This insight comes from an assumption that people base their expectations of inflation on their observation of past inflation. If, instead, people are more forward looking and understand what the policymaker is trying to do, they might adjust their expectations more quickly, causing the rise in inflation to lose much of even its temporary effect on real activity. In a sense, even the short-run relationship relied on people being fooled. One way people might be fooled is if they are simply unable to distinguish general inflation from a change in relative prices. This confusion, sometimes referred to as *money illusion*, could cause people to react to inflation as if it were a change in relative prices. For instance, workers, seeing their nominal wages rise but not recognizing that a general inflation is in process, might react as if their real income were rising. That is, they might increase their expenditures on goods and services.

Robert Lucas, another Nobel Laureate, demonstrated how behavior resembling money illusion could result even with firms and consumers who fully understood the difference between relative prices and the general price level.[9] In his analysis, confusion comes not from people's misunderstanding, but from their inability to observe all of the economy's prices at one time. His was the first formal analysis showing how a Phillips curve relationship could emerge in an economy with forward-looking decisionmakers. Like the work of Friedman and Phelps, Lucas' implications for policymakers were cautionary. The relationship between inflation and real activity in his analysis emerged most strongly when policy was conducted in an unpredictable fashion, that is, when policymaking was more a source of volatility than stability.

The Great Inflation

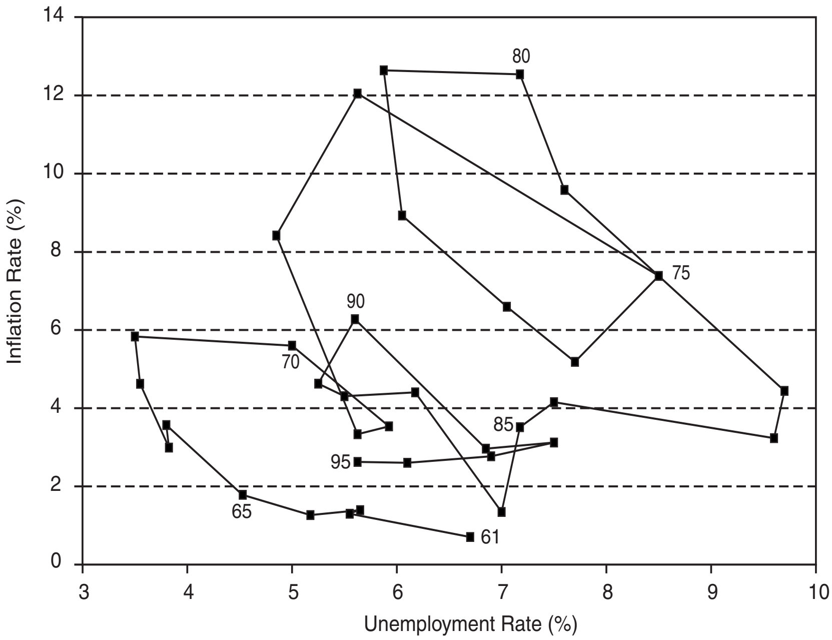
The expectations-augmented Phillips curve had the stark implication that any attempt to utilize the relationship between inflation and real activity to engineer persistently low unemployment at the cost of a little more inflation was doomed to failure. The experience of the 1970s is widely taken to be a confirmation of this hypothesis. The historical relationship identified by Phillips, Samuelson, and Solow, and other earlier writers appeared to break down entirely, as shown by the scatter-plot of the data for the 1970s in Figure 4. Throughout this decade, both inflation and unemployment tended to grow, leading to the emergence of the term "stagflation" in the popular lexicon.

One possible explanation for the experience of the 1970s is that the decade was simply a case of bad luck. The Phillips curve shifted about unpredictably as the economy was battered by various external shocks. The most notable of these shocks were the dramatic increases in energy prices in 1973 and again later in the decade. Such supply shocks worsened the available tradeoff, making higher unemployment necessary at any given level of inflation.

By contrast, viewing the decade through the lens of the expectations-augmented Phillips curve suggests that policy shared the blame for the disappointing results. Policymakers attempted to shield the real economy from the effects of aggregate shocks. Guided by the Phillips curve, this effort often implied a choice to tolerate higher inflation rather than allowing unemployment to rise. This type of policy choice follows from viewing the statistical relationship Phillips first found in the data as a menu of policy options, as suggested by Samuelson and Solow. But the arguments made by Friedman and Phelps imply that such a tradeoff is short lived at best. Unemployment would ultimately return to its natural rate at the higher rate of inflation. So, while the relative importance of luck and policy for the poor macroeconomic perfor-

[9] Lucas (1972).

Figure 4 Inflation-Unemployment Relationship in the United States, 1961–1995



Notes: Inflation rate is seasonally-adjusted CPI, Fourth Quarter.

Source: Bureau of Labor Statistics/Haver Analytics.

mance of the 1970s continues to be debated by economists, we find a powerful lesson in the history of that decade.[10] The macroeconomic performance of the 1970s is largely what the expectations-augmented Phillips curve predicts when policymakers try to exploit a tradeoff that they mistakenly believe to be stable.

The insights of Friedman, Phelps, and Lucas pointed to the complicated interaction between policymaking and statistical analysis. Relationships we observe in past data were influenced by past policy. When policy changes, people's behavior may change and so too may statistical relationships. Hence, the history of the 1970s can be read as an illustration of Lucas' critique of what was at the time the consensus approach to policy analysis.[11]

[10] Velde (2004) provides an excellent overview of this debate. A nontechnical description of the major arguments can be found in Sumo (2007).

[11] Lucas (1976).

Focusing attention on the role of expectations in the Phillips curve creates a challenge for policymakers seeking to use monetary policy to manage real economic activity. At any point in time, the current state of the economy and the private sector's expectations may imply a particular Phillips curve. Assuming that the Phillips curve describes a stable relationship, a policymaker might choose a preferred inflation-unemployment combination. That very choice, however, can alter expectations, causing the tradeoff to change. The policymaker's problem is, in effect, a game played against a public that is trying to anticipate policy. What's more, this game is repeated over and over, each time a policy choice must be made. This complicated interdependence of policy choices and private sector actions and expectations was studied by Finn Kydland and Edward C. Prescott.[12] In one of the papers for which they were awarded the 2005 Nobel Prize, they distinguish between rules and discretion as approaches to policymaking. By discretion, they mean period-by-period decisionmaking in which the policymaker takes a fresh look at the costs and benefits of alternative inflation levels at each moment. They contrast this with a setting in which the policymaker makes a one-time decision about the best rule to guide policy. They show that discretionary policy would result in higher inflation and no lower unemployment than the once-and-for-all choice of a policy rule.

Recent work by Thomas Sargent and various coauthors shows how discretionary policy, as studied by Kydland and Prescott, can lead to the type of inflation outcomes experienced in the 1970s.[13] This analysis assumes that the policymaker is uncertain of the position of the Phillips curve. In the face of this uncertainty, the policymaker estimates a Phillips curve from historical data. Seeking to exploit a short-run, expectations-augmented Phillips curve—that is, pursuing discretionary policy—the policymaker chooses among inflation-unemployment combinations described by the estimated Phillips curve. But the policy choices themselves cause people's beliefs about policy to change, which causes the response to policy choices to change. Consequently, when the policymaker uses new data to update the estimated Phillips curve, the curve will have shifted. This process of making policy while also trying to learn about the location of the Phillips curve can lead a policymaker to choices that result in persistently high inflation outcomes.

In addition to the joint rise in inflation and unemployment during the 1970s, other empirical evidence pointed to the importance of expectations. Sargent studied the experience of countries that had suffered from very high inflation.[14] In countries where monetary reforms brought about sudden and rapid *decelerations* in inflation, he found that the cost in terms of reduced

[12] Kydland and Prescott (1977).

[13] Sargent (1999), Cogley and Sargent (2005), and Sargent, Williams, and Zha (2006).

[14] Sargent (1986).

output or increased unemployment tended to be much lower than standard Phillips curve tradeoffs would suggest. One interpretation of these findings is that the disinflationary policies undertaken tended to be well-anticipated. Policymakers managed to credibly convince the public that they would pursue these policies. Falling inflation that did not come as a surprise did not have large real economic costs.

On a smaller scale in terms of peak inflation rates, another exercise in dramatic disinflation was conducted by the Federal Reserve under Chairman Paul Volcker.[15] As inflation rose to double-digit levels in the late 1970s, contemporaneous estimates of the cost in unemployment and lost output that would be necessary to bring inflation down substantially were quite large. A common range of estimates was that the 6 percentage-point reduction in inflation that was ultimately brought about would require output from 9 to 27 percent below capacity annually for up to four years.[16] Beginning in October 1979, the Fed took drastic steps, raising the federal funds rate as high as 19 percent in 1980. The result was a steep, but short recession. Overall, the costs of the Volcker disinflation appear to have been smaller than had been expected. A standard estimate, which appears in a popular economics textbook, is one in which the reduction in output during the Volcker disinflation amounted to less than a 4 percent annual shortfall relative to capacity.[17] This amount is a significant cost, but it is substantially less than many had predicted before the fact. Again, one possible reason could be that the Fed's course of action in this episode became well-anticipated once it commenced. While the public might not have known the extent of the actions the Fed would take, the direction of the change in policy may well have become widely understood. By the same token, and as argued by Goodfriend and King, remaining uncertainty about how far and how persistently the Fed would bring inflation down may have resulted in the costs of disinflation being greater than they might otherwise have been.

The experience of the 1970s, together with the insights of economists emphasizing expectations, ultimately brought the *credibility* of monetary policy to the forefront in thinking about the relationship between inflation and the real economy. Credibility refers to the extent to which the central bank can convince the public of its intention with regard to inflation. Kydland and Prescott showed that credibility does not come for free. There is always a short-run gain from allowing inflation to rise a little so as to stimulate the real economy. To establish credibility for a low rate of inflation, the central bank must convince the public that it will not pursue that short-run gain.

---

[15] Goodfriend and King (2005).

[16] Ibid.

[17] Mankiw (2007).

The experience of the 1980s and 1990s can be read as an exercise in building credibility. In several episodes during that period, inflation expectations rose as doubts were raised about the Fed's ability to maintain its commitment to low inflation. These episodes, labeled *inflation scares* by Marvin Goodfriend, were marked by rapidly rising spreads between long-term and short-term interest rates.[18] Goodfriend identifies inflation scares in 1980, 1983, and 1987. These tended to come during or following episodes in which the Fed responded to real economic weakness with reductions (or delayed increases) in its federal funds rate target. In these instances, Fed policymakers reacted to signs of rising inflation expectations by raising interest rates. These systematic policy responses in the 1980s and 1990s were an important part of the process of building credibility for lower inflation.

2. THE "MODERN" PHILLIPS CURVE

The history of the Phillips curve shows that the empirical relationship shifts over time, and there is evidence that those movements are linked to the public's inflation expectations. But what does the history say about why this relationship exists? Why is it that there is a statistical relationship between inflation and real economic activity, even in the short run? The earliest writers and those that followed them recognized that the short-run tradeoff must arise from frictions that stand in the way of monetary neutrality. There are many possible sources of such frictions. They may arise from the limited nature of the information individuals have about the full array of prices for all products in the economy, as emphasized by Lucas. Frictions might also stem from the fact that not all people participate in all markets, so that different markets might be affected differently by changes in monetary policy. One simple type of friction is a limitation on the flexibility sellers have in adjusting the prices of the goods they sell. If there are no limitations all prices can adjust seamlessly whenever demand or cost conditions change, then a change in monetary policy will, again, affect different markets differently.

Deriving a Phillips Curve from Price-Setting Behavior

This price-setting friction has become a popular device for economists seeking to model the behavior of economies with a short-run Phillips curve. To see how such a friction leads to a Phillips curve, think about a business that is setting a price for its product and does not expect to get around to setting the price again for some time. Typically, the business will choose a price based on its own costs of production and the demand that it faces for its goods. But

[18] Goodfriend (1993).

because that business expects its price to be fixed for a while, its price choice will also depend on what it expects to happen to its costs and its demand between when it sets its price this time and when it sets its price the next time.

If the price-setting business thinks that inflation will be high in the interim between its price adjustments, then it will expect its relative price to fall. As average prices continue to rise, a good with a temporarily fixed price gets cheaper. The firm will naturally be interested in its average relative price during the period that its price remains fixed. The higher the inflation expected by the firm up until its next price adjustment, the higher the current price it will set. This reasoning, applied to all the economy's sellers of goods and services, leads directly to a close relationship between current inflation and expected future inflation.

This description of price-setting behavior implies that current inflation depends on the real costs of production and expected future inflation. The real costs of production for businesses will rise when the aggregate use of productive resources rises, for instance because rising demand for labor pushes up real wages.[19] The result is a Phillips curve relationship between inflation and a measure of real economic activity, such as output growth or unemployment. Current inflation rises with expected future inflation and falls as current unemployment rises relative to its "natural" rate (or as current output falls relative to the trend rate of output growth).

A Phillips Curve in a "Complete" Modern Model

The price-setting frictions that are part of many modern macroeconomic models are really not that different from arguments that economists have always made about reasons for the short-run nonneutrality of money. What distinguishes the modern approach is not just the more formal, mathematical derivation of a Phillips curve relationship, but more importantly, the incorporation of this relationship into a complete model of the macroeconomy. The word "complete" here has a very specific meaning, referring to what economists call "general equilibrium." The general equilibrium approach to studying economic activity recognizes the interdependence of disparate parts of the economy and emphasizes that all macroeconomic variables such as GDP, the level of prices, and unemployment are all determined by fundamental economic forces acting at the level of individual households and businesses. The completeness of a general equilibrium model also allows for an analysis of the effects of alternative approaches to macroeconomic policy, as well as an evaluation of the relative merits of alternative policies in terms of their effects on the economic well-being of the people in the economy.

---

[19] There are a number of technical assumptions needed to make this intuitive connection precisely correct.

The Phillips curve is only one part of a complete macroeconomic model—one equation in a system of equations. Another key component describes how real economic activity depends on real interest rates. Just as the Phillips curve is derived from a description of the price-setting decisions of businesses, this other relationship, which describes the demand side of the economy, is based on households' and business' decisions about consumption and investment. These decisions involve people's demand for resources now, as compared to their expected demand in the future. Their willingness to trade off between the present and the future depends on the price of that tradeoff—the real rate of interest.

One source of interdependence between different parts of the model—different equations—is in the real rate of interest. A real rate is a nominal rate—the interest rates we actually observe in financial markets—adjusted for expected inflation. Real rates are what really matter for households' and firms' decisions. So on the demand side of the economy, people's choices about consumption and investment depend on what they expect for inflation, which comes, in part, from the pricing behavior described by the Phillips curve. Another source of interdependence comes in the way the central bank influences nominal interest rates by setting the rate charged on overnight, interbank loans (the federal funds rate in the United States). A complete model also requires a description of how the central bank changes its nominal interest rate target in response to changing economic conditions (such as inflation, growth, or unemployment).

In a complete general equilibrium analysis of an economy's performance, all three parts—the Phillips curve, the demand side, and central bank behavior—work together to determine the evolution of economic variables. But many of the economic choices people make on a day-to-day basis depend not only on conditions today, but also on how conditions are expected to change in the future. Such expectations in modern macroeconomic models are commonly described through the assumption of *rational expectations*. This assumption simply means that the public—households and firms whose decisions drive real economic activity—fully understands how the economy evolves over time and how monetary policy shapes that evolution. It also means that people's decisions will depend on well-informed expectations not only of the evolution of future fundamental conditions, but of future policy as well. While discussions of a central bank's credibility typically assume that there are things related to policymaking about which the public is not fully certain, these discussions retain the presumption that people are forward looking in trying to understand policy and its impact on their decisions.

Implications and Uses of the Modern Approach

A Phillips curve that is derived as part of a model that includes price-setting frictions is often referred to as the *New Keynesian Phillips curve* (NKPC).[20] A complete general equilibrium model that incorporates this version of the Phillips curve has been referred to as the *New Neoclassical Synthesis model*.[21] These models, like any economic model, are parsimonious descriptions of reality. We do not take them as exact descriptions of how a modern economy functions. Rather, we look to them to capture the most important forces at work in determining macroeconomic outcomes. The key equations in new neoclassical or new Keynesian models all involve assumptions or approximations that simplify the analysis without altering the fundamental economic forces at work. Such simplifications allow the models to be a useful guide to our thinking about the economy and the effects of policy.

The modern Phillips curve is similar to the expectations-augmented Phillips curve in that inflation expectations are important to the relationship between current inflation and unemployment. But its derivation from forward-looking price-setting behavior shifts the emphasis to expectations of future inflation. It has implications similar to the long-run neutrality of money, because if inflation is constant over time, then current inflation is equal to expected inflation. Then, whatever that constant rate of inflation, unemployment must return to the rate implied by the underlying structure of the economy, that is, to a rate that might be considered the "natural" unemployment. Money is not truly neutral in these models, however. Rather, the pricing frictions underlying the models imply that there are real economic costs to inflation. Because sellers of goods adjust their prices at different times, inflation makes the relative prices of different goods vary, and this distorts sellers' and buyers' decisions. This distortion is greater, the greater the rate of inflation.

The expectational nature of the Phillips curve also means that policies that have a short-run effect on inflation will induce real movements in output or unemployment mainly if the short-run movement in inflation is not expected to persist. In this sense, the modern Phillips curve also embodies the importance of monetary policy credibility, since it is credibility that would allow expected inflation to remain stable, even as inflation fluctuated in the near term.

A more general way of emphasizing the importance of credibility is to say that the modern Phillips curve implies that the behavior of inflation will depend crucially on people's understanding of how the central bank is conducting monetary policy. What people think about the central bank's objectives and strategy will determine expectations of inflation, especially over the long run. Uncertainty about these aspects of policy will cause people to try to make

[20] Clarida, Galí, and Gertler (1999).

[21] Goodfriend and King (1997).

inferences about future policy from the actual policy they observe. Even if the central bank makes statements about its long-run objectives and strategy, people will still try to make inferences from the policy actions they see. But in this case, the inference that people will try to make is slightly simpler: people must determine if actual policy is consistent with the stated objectives.

Does this newest incarnation of the Phillips curve present a central bank with the opportunity to actively manage real economic activity through choosing more or less inflationary policies? The assumption that people are forward looking in forming expectations about future policy and inflation limits the scope for managing real growth or unemployment through Phillips curve trade-offs. An attempt to manage such growth or unemployment persistently would translate into the public's expectations of inflation causing the Phillips curve to shift. This is another characteristic that the modern approach shares with the older expectations-augmented Phillips curve.

What this modern framework does allow is the analysis of alternative monetary policy rules—that is, how the central bank sets its nominal interest rate in response to such economic variables as inflation, relative to the central bank's target, and the unemployment rate or the rate of output growth relative to the central bank's understanding of trend growth.[22] A typical rule that roughly captures the actual behavior of most central banks would state, for instance, that the central bank raises the interest rate when inflation is higher than its target and lowers the interest rate when unemployment rises. Alternative rules might make different assumptions, for instance, about *how much* the central bank moves the interest rate in response to changes in the macroeconomic variables that it is concerned about. The complete model can then be used to evaluate how different rules perform in terms of the long-run levels of inflation and unemployment they produce, or more generally in terms of the economic well-being generated for people in the economy. A typical result is that rules that deliver lower and less variable inflation are better both because low and stable inflation is a good thing and because such rules can also deliver less variability in real economic activity. Further, lower inflation has the benefit of reducing the costs from distorted relative prices.

While low inflation is a preferred outcome, it is typically not possible, in models or in reality, to engineer a policy that delivers the same low target rate of inflation every month or quarter. The economy is hit by any number of shocks that can move both real output and inflation around from month to month—large energy price movements, for example. In the presence of such shocks, a good policy might be one that, while not hitting its inflation target each month, always tends to move back toward its target and never stray too far.

[22] We use the term "monetary policy rule" in the very general sense of any systematic pattern of choice for the policy instrument—the funds rate—based on the state of the economy.

Complete models incorporating a modern Phillips curve also allow economists to formalize the notion of monetary policy credibility. Remember that credibility refers to what people believe about the way the central bank intends to conduct policy. If people are uncertain about what rule best describes the behavior of the central bank, then they will try to learn from what they see the central bank doing. This learning can make people's expectations about future policy evolve in a complicated way. In general, uncertainty about the central bank's policy, or doubts about its commitment to low inflation, can raise the cost (in terms of output or employment) of reducing inflation. That is, the short-run relationship between inflation and unemployment depends on the public's long-run expectations about monetary policy and inflation.

The modern approach embodies many features of the earlier thinking about the Phillips curve. The characterization of policy as a systematic pattern of behavior employed by the central bank, providing the framework within which people form systematic expectations about future policy, follows the work of Kydland and Prescott. And the focus on expectations itself, of course, originated with Friedman. Within this modern framework, however, some important debates remain unsettled. While our characterization of the framework has emphasized the forward-looking nature of people's expectations, some economists believe that deviations from this benchmark are important for understanding the dynamic behavior of inflation. We turn to this question in the next section.

We have described here an approach that has been adopted by many contemporary economists for applied central bank policy analysis. But we should note that this approach is not without its critics. Many economists view the price-setting frictions that are at the core of this approach as ad hoc and unpersuasive. This critique points to the value of a deeper theory of firms' price-setting behavior. Moreover, there are alternative frictions that can also rationalize monetary nonneutrality. Alternatives include frictions that limit the information available to decisionmakers or that limit some people's participation in some markets. So while the approach we've described does not represent the only possible modern model, it has become a popular workhorse in policy research.

3. HOW WELL DOES THE MODERN PHILLIPS CURVE FIT THE DATA?

The Phillips curve began as a relationship drawn to fit the data. Over time, it has evolved as economists' understanding of the forces driving those data has developed. The interplay between theory—the application of economic logic—and empirical facts has been an important part of this process of discovery. The recognition of the importance of expectations developed together with the evidence of the apparent instability of the short-run tradeoff. The modern Phillips curve represents an attempt to study the behavior of both inflation and real variables using models that incorporate the lessons of Friedman, Phelps, and Lucas and that are rich enough to produce results that can be compared to real world data.

Attempts to fit the modern, or New Keynesian, Phillips curve to the data have come up against a challenging finding. The theory behind the short-run relationship implies that current inflation should depend on current real activity, as measured by unemployment or some other real variable, and expected future inflation. When estimating such an equation, economists have often found that an additional variable is necessary to explain the behavior of inflation over time. In particular, these studies find that past inflation is also important.[23]

Inflation Persistence

The finding that past inflation is important for the behavior of current and future inflation—that is, the finding of inflation *persistence*—implies that movements in inflation have persistent effects on future inflation, apart from any effects on unemployment or expected inflation. Such persistence, if it were an inherent part of the structure and dynamics of the economy, would create a challenge for policymakers to reduce inflation by reducing people's expectations. Remember that we stated earlier the possibility that if the central bank could convince the public that it was going to bring inflation down, then the desired reduction might be achieved with little cost in unemployment or output. Inherent inflation persistence would make such a strategy problematic. Inherent persistence makes the set of choices faced by the policymaker closer to that originally envisioned by Samuelson and Solow. The faster one tries to bring down inflation, the greater the real economic costs.

Inherent persistence in inflation might be thought to arise if not all pricesetters in the economy were as forward looking as in the description given earlier. If, instead of basing their price decisions on their best forecast of future inflation behavior, some firms simply based current price choices on the past behavior of inflation, this backward-looking pricing would impart persistence to inflation. Jordi Galí and Mark Gertler, who took into account the possibility that the economy is populated by a combination of forward-looking and backward-looking participants, introduced a *hybrid* Phillips curve in which current inflation depends on both expected future inflation and past inflation.[24]

[23] Fuhrer (1997).

[24] Galí and Gertler (1999).

An alternative explanation for inflation persistence is that it is a result primarily of the conduct of monetary policy. The evolution of people's inflation expectations depends on the evolution of the conduct of policy. If there are significant and persistent shifts in policy conduct, expectations will evolve as people learn about the changes. In this explanation, inflation persistence is not the result of backward-looking decisionmakers in the economy but is instead the result of the interaction of changing policy behavior and forward-looking private decisions by households and businesses.[25]

Another possibility is that inflation persistence is the result of the nature of the shocks hitting the economy. If these shocks are themselves persistent—that is, bad shocks tend to be followed by more bad shocks—then that persistence can lead to persistence in inflation. The way to assess the relative importance of alternative possible sources of persistence is to estimate the multiple equations that make up a more complete model of the economy. This approach, in contrast with the estimation of a single Phillips curve equation, allows for explicitly considering the roles of changing monetary policy, backward-looking pricing behavior, and shocks in generating inflation persistence. A typical finding is that the backward-looking terms in the hybrid Phillips curve appear considerably less important for explaining the dynamics of inflation than in single equation estimation.[26]

The scientific debate on the short-run relationship between inflation and real economic activity has not yet been fully resolved. On the central question of the importance of backward-looking behavior, common sense suggests that there are certainly people in the real-world economy who behave that way. Not everyone stays up-to-date enough on economic conditions to make sophisticated, forward-looking decisions. People who do not may well resort to rules of thumb that resemble the backward-looking behavior in some economic models. On the other hand, people's behavior is bound to be affected by what they believe to be the prevailing rate of inflation. Market participants have ample incentive and ability to anticipate the likely direction of change in the economy. So both backward- and forward-looking behavior are grounded in common sense. However the more important scientific questions involve the extent to which either type of behavior drives the dynamics of inflation and is therefore important for thinking about the consequences of alternative policy choices.

---

[25] Dotsey (2002) and Sbordone (2006).

[26] Lubik and Schorfheide (2004).

The Importance of Inflation Persistence for Policymakers

Related to the question of whether forward- or backward-looking behavior drives inflation dynamics is the question of how stable people's inflation expectations are. The backward-looking characterization suggests a stickiness in beliefs, implying that it would be hard to induce people to change their expectations. If relatively high inflation expectations become ingrained, then it would be difficult to get people to expect a decline in inflation. This describes a situation in which disinflation could be very costly, since only persistent evidence of changes in actual inflation would move future expectations. Evidence discussed earlier from episodes of dramatic changes in the conduct of policy, however, suggests that people can be convinced that policy has changed. In a sense, the tradeoffs faced by a policymaker could depend on the extent to which people's expectations are subject to change. If people are uncertain and actively seeking to learn about the central bank's approach to policy, then expectations might move around in a way that departs from the very persistent, backward-looking characterization. But this movement in expectations would depend on the central bank's actions and statements about its conduct of policy.

The periods that Goodfriend (1993) described as inflation scares can be seen as periods when people's assessment of likely future policy was changing rather fluidly. Even very recently, we have seen episodes that could be described as "mini scares." For instance, in the wake of Hurricane Katrina in late 2005, markets' immediate response to rising energy prices suggested expectations of persistently rising inflation. Market participants, it seems, were uncertain as to how much of a run-up in general inflation the Fed would allow. Inflation expectations moved back down after a number of FOMC members made speeches emphasizing their focus on preserving low inflation. This episode illustrates both the potential for the Fed to influence inflation expectations and the extent to which market participants are at times uncertain as to how the Fed will respond to new developments.

4. MAKING POLICY

While the scientific dialogue continues, policymakers must make judgments based on their understanding of the state of the debate. At the Federal Reserve Bank of Richmond, policy opinions and recommendations have long been guided by a view that the short-term costs of reducing inflation depend on expectations. This view implies that central bank credibility—that is, the public's level of confidence about the central bank's future patterns of behavior—is an important aspect of policymaking. Central bank credibility makes it less costly to return inflation to a desirable level after it has been pushed up (or down) by energy prices or other shocks to the economy. This view of policy is consistent with a view of the Phillips curve in which inflation persistence is primarily a consequence of the conduct of policy.

The evidence is perhaps not yet definitive. As outlined in our argument, however, we do find support for our view in the broad contours of the history of U.S. inflation over the last several decades. At a time when a consensus developed in the economics profession that the Phillips curve tradeoff could be exploited by policymakers, apparent attempts to do so led to or contributed to the decidedly unsatisfactory economic performance of the 1970s. And the improved performance that followed coincided with the solidification of the profession's understanding of the role of expectations. We also see the initial costs of bringing down inflation in the early 1980s as consistent with our emphasis on expectations and credibility. After the experience of the 1970s, credibility was low, and expectations responded slowly to the Fed's disinflationary policy actions. Still, the response of expectations was faster than might be implied by a backward-looking Phillips curve.

We also view policymaking on the basis of a forward-looking understanding of the Phillips curve as a prudent approach. A hybrid Phillips curve with a backward-looking component presents greater opportunities for exploiting the short-run tradeoff. In a sense, it assumes that the monetary policymaker has more influence over real economic activity than is assumed by the purely forward-looking specification. Basing policy on a backward-looking formulation would also risk underestimating the extent to which movements in inflation can generate shifts in inflation expectations, which could work against the policymaker's intentions. Again, the experience of past decades suggests the risks associated with policymaking under the assumption that policy can persistently influence real activity more than it really can. In our view, these risks point to the importance of a policy that makes expectational stability its centerpiece.

5. CONCLUSION

One key lesson from the history of the relationship between inflation and real activity is that any short-run tradeoff depends on people's expectations for inflation. Ultimately, monetary policy has its greatest impact on real activity when it deviates from people's expectations. But if a central bank tries to deviate from people's expectations repeatedly, so as to systematically increase real output growth, people's expectations will adjust.

There are also, we think, important lessons in the observation that overall economic performance, in terms of both real economic activity and inflation, was much improved beginning in the 1980s as compared to that in the preceding decade. While this improvement could have some external sources related to the kinds of shocks that affect the economy, it is also likely that improved conduct of monetary policy played a role. In particular, monetary policy was able to persistently lower inflation by responding more to signs of rising inflation or inflation expectations than had been the case in the past. At the same time, the variability of inflation fell, while fluctuations in output and unemployment were also moderating.

We think the observed behavior of policy and economic performance is directly linked to the lessons from the history of the Phillips curve. Both point to the importance of the expectational consequences of monetary policy choices. An approach to policy that is able to stabilize expectations will be most able to maintain low and stable inflation with minimal effects on real activity. It is the credible maintenance of price stability that will in turn allow real economic performance to achieve its potential over the long run. This will not eliminate the business cycle since the economy will still be subject to shocks that quicken or slow growth. We believe the history of the Phillips curve shows that monetary policy's ability to add to economic variability by overreacting to shocks is greater than its ability to reduce real variability, once it has achieved credibility for low inflation.

REFERENCES

Clarida, Richard, Jordi Galí, and Mark Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37 (4): 1,661–707.

Cogley, Timothy, and Thomas J. Sargent. 2005. "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty." *Review of Economic Dynamics* 8 (2): 528–63.

Cooley, Thomas F., and Gary D. Hansen. 1989. "The Inflation Tax in a Real Business Cycle Model." *American Economic Review* 79 (4): 733–48.

Dotsey, Michael. 2002. "Structure from Shocks." Federal Reserve Bank of Richmond *Economic Quarterly* 88 (4): 37–47.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58 (1): 1–17.

Fuhrer, Jeffrey C. 1997. "The (Un)Importance of Forward-Looking Behavior in Price Specifications." *Journal of Money, Credit, and Banking* 29 (3): 338–50.

Galí, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2): 195–222.

Goodfriend, Marvin. 1993. "Interest Rate Policy and the Inflation Scare Problem: 1979-1992." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (1): 1–24.

Goodfriend, Marvin, and Robert G. King. 1997. "The New Neoclassical Synthesis and the Role of Monetary Policy." In *NBER Macroeconomics Annual 1997*, eds. Ben S. Bernanke and Julio J. Rotemberg. Cambridge, MA: The MIT Press.

Goodfriend, Marvin, and Robert G. King. 2005. "The Incredible Volcker Disinflation." *Journal of Monetary Economics* 52 (5): 981–1,015.

Humphrey, Thomas M. 1991. "Nonneutrality of Money in Classical Monetary Thought." Federal Reserve Bank of Richmond *Economic Review* 77 (2): 3–15.

Keynes, John Maynard. 1923. *A Tract on Monetary Reform*. London: Macmillan and Company.

King, Robert G., and Charles I. Plosser. 1984. "Money, Credit, and Prices in a Real Business Cycle." *American Economic Review* 74 (3): 363–80.

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

Lubik, Thomas A., and Frank Schorfheide. 2004. "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94 (1): 190–217.

Lucas, Robert E., Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4 (2): 103–24.

Lucas, Robert E., Jr. 1976. "Econometric Policy Evaluation: A Critique." In the Carnegie-Rochester Conference Series on Public Policy 1, *The Phillips Curve and Labor Markets*, eds. Karl Brunner and Allan H. Meltzer. Amsterdam: North-Holland.

Lucas, Robert E., Jr. 1996. "Nobel Lecture: Monetary Neutrality." *Journal of Political Economy* 104 (4): 661–82.

Mankiw, N. Gregory. 2007. *Principles of Microeconomics* (4th ed.). United States: Thomson South-Western.

Phelps, Edmund S. 1967. "Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time." *Economica* 34 (135): 254–81.

Phillips, A. W. 1958. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957." *Economica* 25 (100): 283–99.

Samuelson, Paul A., and Robert M. Solow. 1960. "Analytical Aspects of Anti-Inflation Policy." *American Economic Review* 50 (2): 177–94.

Sargent, Thomas J. 1986. "The Ends of Four Big Inflations." In *Rational Expectations and Inflation*, ed. Thomas J. Sargent. New York, NY: Harper and Row.

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

Sargent, Thomas, Noah Williams, and Tao Zha. 2006. "Shocks and Government Beliefs: The Rise and Fall of American Inflation." *American Economic Review* 96 (4): 1,193–224.

Sbordone, Argia M. 2006. "Inflation Persistence: Alternative Interpretations and Policy Implications." Federal Reserve Bank of New York, Manuscript (October).

Sumo, Vanessa. 2007. "Bad Luck or Bad Policy? Why Inflation Rose and Fell, and What This Means for Monetary Policy." Federal Reserve Bank of Richmond *Region Focus* 11 (1): 40–3.

Velde, François R. 2004. "Poor Hand or Poor Play? The Rise and Fall of Inflation in the U.S." Federal Reserve Bank of Chicago *Economic Perspectives* 28 (1): 34–51.

Economic Quarterly—Volume 93, Number 3—Summer 2007—Pages 229–250

A Taylor Rule and the Greenspan Era

Yash P. Mehra and Brian D. Minton

There is considerable interest in determining whether monetary policy actions taken by the Federal Reserve under Chairman Alan Greenspan can be summarized by a Taylor rule. The original Taylor rule relates the federal funds rate target to two economic variables: lagged inflation and the output gap, with the actual federal funds rate completely adjusting to the target in each period (Taylor 1993).[1] The later assumption of complete adjustment has often been interpreted as indicating the policy rule is "non-inertial," or the Federal Reserve does not smooth interest rates. Inflation in the original Taylor rule is measured by the behavior of the GDP deflator and the output gap is the deviation of the log of real output from a linear trend. Taylor (1993) shows that from 1987 to 1992 policy actions did not differ significantly from prescriptions of this simple rule. Hence, according to the original Taylor rule, the Federal Reserve, at least during the early part of the Greenspan era, was backward looking, focused on headline inflation, and followed a non-inertial policy rule.

Recent research, however, suggests a different picture of the Federal Reserve under Chairman Greenspan. English, Nelson, and Sack (2002) present evidence that indicates policy actions during the Greenspan period are better explained by an "inertial" Taylor rule reflecting the presence of interest rate smoothing.[2] Blinder and Reis (2005) state that the Greenspan Fed focused on

We would like to thank Andreas Hornstein, Robert Hetzel, Roy Webb, and Nashat Moin for their comments. The views expressed in this article are those of the authors and do no necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Taylor (1993) did not estimate the policy rule but chose specific values for the policy response coefficients, the real rate, and the inflation target.

[2] English, Nelson, and Sack (2002) provide empirical evidence for the hypothesis that the Greenspan Fed smoothed interest rates. Woodford (2005) suggests the Federal Reserve under Greenspan, in fact, communicated its interest-smoothing intentions to financial markets by including descriptive, forward-looking sentences in its policy statements to ensure that policy expectations of the financial sector remain aligned with its own outlook for policy. For example, in order to

a "core" measure of inflation in adjusting its federal funds rate target. Clarida, Galí, and Gertler (2000), among others, have shown that a forward-looking Taylor rule that relates the current funds rate target to "expected" inflation and output developments appears to fit the data quite well over the period spanning the tenures of Chairmen Paul Volcker and Alan Greenspan. Orphanides (2001) argues that policy evaluations using policy rules estimated with the final revised data may be misleading.

This article estimates a Taylor rule that address three key features of the Greenspan period highlighted in recent research: the Federal Reserve under Greenspan was forward looking, focused on core inflation, and smoothed interest rates. Furthermore, this article uses the real-time data for economic variables and investigates whether results based on the final, revised data change when the real-time data are used. We also examine whether the use of real-time data leads to a better explanation of policy actions during the Greenspan period.

A Taylor rule incorporating the above-noted three features is shown below in equation (1.3).

$$\begin{aligned} FR_t^* &= \alpha_0 + \alpha_\pi \pi_{t,j}^c + \alpha_y (ln\ y_{t,k} - \ ln\ y_{t,k}^*), && (1.1) \\ FR_t &= \rho FR_{t-1} + (1-\rho) FR_t^* + v_t, && (1.2) \\ FR_t &= \rho FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi \pi_{t,j}^c && (1.3) \\ & \quad + \alpha_y (ln\ y_{t,k} - ln y_{t,k}^*)\} + v_t, \end{aligned}$$

where FR_t is the actual federal funds rate, FR_t^* is the federal funds rate target, $\pi_{t,j}^c$ is the j-period ahead forecast of core inflation made at time t, $ln\ y$ is the log of actual output, $ln\ y^*$ is the log of potential output, and v_t is the disturbance term. Thus, the term $(ln\ y_{t,k} - \ ln\ y_{t,k}^*)$ is the k-period ahead forecast of the output gap. Equation (1.1) relates the federal funds rate target to expected values of two economic fundamentals: core inflation and the output gap. The funds rate target is hereafter called the policy rate. The coefficients α_π and α_y measure the long-term responses of the funds rate target to the expected inflation and the output gap. They are assumed to be positively signed, indicating that the Federal Reserve raises its funds rate

deal with the threat of deflation in 2003, policy statements in that year included sentences such as "...policy accommodation can be maintained for a considerable period of time," meaning the Federal Reserve would not raise its funds rate target in response to increases in real growth given the threat of deflation. The intent was to hold long-term interest rates low by quashing expectations that the Fed was on the verge of increasing the funds rate. In 2004, policy statements included phrases such as "...the Committee believes that it can be patient in removing policy accommodation," and "...the Committee believes that policy accommodation can be removed at a pace that is likely to be measured." The latter came to mean 25 basis points at each FOMC meeting. These considerations suggest the Greenspan policy rule should be estimated allowing for the presence of interest-rate smoothing. Blinder and Reis (2005) also argue that the Greenspan Fed used frequent small changes in the funds rate to hit its target for the policy rate suggested by economic fundamentals such as inflation and unemployment.

target if inflation rises and/or the output gap is positive. Equation (1.2) is the standard partial adjustment equation, expressing the current funds rate as a weighted average of the current funds rate target FR_t^* and last quarter's actual value FR_{t-1}. If the actual funds rate adjusts to its target within each period, then ρ equals zero, which suggests that the Federal Reserve does not smooth interest rates. Equation (1.2) also includes a disturbance term, indicating that in the short run, the actual funds rate may deviate from the value implied by economic determinants specified in the policy rule. If we substitute equation (1.1) into (1.2), we get (1.3), a forward-looking "inertial" Taylor rule.[3]

This article estimates the Taylor rule (1.3) using final as well as real-time data. The real-time data consists of the Greenbook forecasts of core CPI inflation and the Congressional Budget Office (CBO) estimates of the output gap.[4] The policy rule estimated using the final data covers all of the Greenspan period from 1987:1 to 2005:4, whereas the rule estimated using the Greenbook forecasts spans part of the Greenspan period from 1987:1 to 2000:4, given the five-year lag in release of the Greenbook forecasts to the public.[5]

The empirical work presented here suggests several conclusions. First, policy response coefficients in the estimated inertial Taylor rule (α_π, α_y, ρ) are all positively signed and statistically significant. The key points to note are: (a) the estimated long-term inflation response coefficient α_π is well above unity, which suggests that the Greenspan Fed responded strongly to expected inflation; (b) the estimated output gap response coefficient α_y is generally below unity, suggesting the presence of a relatively weak response to the output gap; and (c) the estimated partial adjustment coefficient ρ is well above zero, indicating the presence of interest-rate smoothing. The conclusion suggested by the estimated Taylor rule, namely, the Greenspan Fed responded strongly to expected inflation developments ($\alpha_\pi > 1$) but relatively weakly to the output gap ($\alpha_y < 1$), is in line with the recent work by Boivin (2006), who, using a different estimation methodology, reports time-varying estimates of inflation and the output gap response coefficients from 1970 to 1995. For the period

---

[3] As is well known, the constant term in the Taylor rule has embedded in it the Federal Reserve's estimates of the short-term real rate and the inflation target. For further explantion, rewrite equation (1.1) of the text as $FR_t^* = rr^* + \pi^* + \alpha_\pi(\pi^c_{t,j} - \pi^*) + \alpha_y(ln\ \ y_{t,k} - ln\ \ y^*_{t,k})$ where $rr^*$is the real rate and π^* is the inflation target. If we substitute the above equation into equation (1.2) of the text, we get equation (1.3) of the text, where the constant term is now defined as $\alpha_0 = rr^* + (1 - \alpha_\pi)\pi^*$. However, one cannot recover estimates of both $rr^*$and π^*without bringing some additional information. See footnote 17.

[4] The preferred measure of real economic activity (say, the output gap) should be the one used in generating the Greenbook forecasts. However, for a major part of the sample period covered here, the Greenbook has not published estimates of the output gap. Hence, it is quite common in this literature to estimate the policy rules using the CBO estimates of the output (or unemployment) gap.

[5] We lose observations at the beginning and end of the sample period due to leads and lags of inflation in the policy rule. The effective sample period is 1988:1 to 2004:4.

since the mid-1980s, the reported estimated policy coefficients are stable and close to values as reported in this article.[6]

Second, the hypothesis that the Greenspan Fed paid attention to expected inflation and output gap developments is supported by additional test results. Those tests favor a forward-looking inertial Taylor rule over the one in which the Federal Reserve focuses on lagged inflation and the output gap. Furthermore, the results somewhat support the hypothesis that the Greenspan Fed was focused on core rather than on headline inflation.

Third, the Taylor rule estimated using the Greenbook core CPI inflation forecasts and the CBO's estimates of real-time output gap has a lower standard error of estimate and predicts policy actions better than the Taylor rule estimated using actual future inflation and the final, revised data on the output gap. However, there still remain several periods during which policy actions differ significantly from prescriptions of the simple Taylor rule. Hence, despite its better fit, the forward-looking inertial Taylor rule estimated here may not be considered a complete description of policy actions taken by the Greenspan Fed.

The rest of the article is organized as follows. Section 1 discusses estimation of the Greenspan policy rule and the real-time data that underlie the estimated policy rule. Section 2 discusses estimation results, and concluding observations are in Section 3.

1. EMPIRICAL METHODOLOGY

Estimation of the Forward-Looking Inertial Taylor Rule

One key objective of this article is to investigate whether monetary policy actions taken by the Federal Reserve under Chairman Greenspan can be summarized by a Taylor rule according to which the Federal Reserve was forward looking, focused on core inflation, and smoothed interest rates. We model the forward-looking nature of the policy rule by relating the current value of the funds rate target to the four-quarter-average expected inflation rate and the contemporaneous output gap. The policy rule incorporating these features is reproduced below in equation (2.3).

[6] In Boivin (2006), the main objective is to investigate whether policy coefficients have changed over time. For expected inflation, the Greenbook forecasts of GNP and GDP deflator are employed. The level of economic activity is proxied using the difference between the natural unemployment rate and the Greenbook forecast of the unemployment rate. The article, however, also uses the real-time output gap measure constructed by Orphanides (2001). For the period 1985 to 1995, the point estimates of the long-run inflation response coefficients are well above unity and those for the long-run output gap response coefficient are well below unity.

$$
\begin{aligned}
FR_t &= \rho\, FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t,\overline{4}} + \alpha_y (ln\ y_t - ln\ y_t^*)\}, \\
&\quad + v_t
\end{aligned} \tag{2.3}
$$

where $\overline{\pi}^c_{t,\overline{4}}$ is the average of one-to-four-quarter-ahead forecasts of core CPI inflation made at time t and other variables as previously defined.[7]

The estimation of the policy rule in equation (2.3) raises several issues. The first issue relates to how we measure expected inflation and the output gap. The second issue relates to the nature of data used in estimation, namely, whether it is the real-time or final, revised data. As discussed earlier, the use of revised as opposed to the real-time data may affect estimates of policy coefficients and may provide a misleading historical analysis of policy actions (Orphanides 2001, 2002). The third issue is an econometric one, arising as a result of the potential presence of serial correlation in the error term v_t. Rudebusch (2006) points out that the Federal Reserve may respond to other economic factors besides expected inflation and the output gap, and hence a Taylor rule estimated omitting those other factors is likely to have a serially correlated error term. The presence of serial correlation in the disturbance term, if ignored, may spuriously indicate that the Federal Reserve is smoothing interest rates.

To further explain that a serially correlated disturbance term may mistakenly indicate the presence of partial adjustment, note first that if the funds rate does partially adjust to the policy rate as shown in (1.2) and the disturbance term has no serial correlation, then the reduced-form policy rule in (1.3 or 2.3) has the lagged funds rate as one of the explanatory variables. Hence, the empirical finding of a significant coefficient on the lagged funds rate in the estimated policy rule may be interpreted as indicating the presence of interest-rate smoothing. Now assume that there is no partial-adjustment, $\rho = 0$ in (2.3), but instead the disturbance term is serially correlated as shown below in equation (3.1).

$$
v_t = s v_{t-1} + \varepsilon_t, \tag{3.1}
$$

$$
\begin{aligned}
FR_t &= sFR_{t-1} + \{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t,\overline{4}} + \alpha_y(\ln y_t - \ln y_t^*)\} \\
&\quad - s\{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t-1,\overline{4}} + \alpha_y(\ln y_{t-1} - \ln y_{t-1}^*)\} + \varepsilon_t.
\end{aligned} \tag{3.2}
$$

If we substitute equation (3.1) into (2.3), it can be easily shown that we get the reduced-form policy rule in equation (3.2), in which among other variables lagged funds rate also enters the policy rule. Hence, the empirical finding of

[7] In particular, the four-quarter-average inflation forecast is defined as $\pi^c_{t,\overline{4}} = \frac{(\pi^c_{t,1}+\pi^c_{t,2}+\pi^c_{t,3}+\pi^c_{t,4})}{4}$. We have also dropped the subscript 0 in the output gap term ($ln\ y_{t,0} - ln\ y^*_{t,0}$).

a significant coefficient on the lagged funds rate in the estimated policy rule may be interpreted arising as a result of interest rate smoothing when in fact, it is not present. In view of these considerations, this policy rule is estimated allowing for the presence of both interest rate smoothing and serial correlation, namely, we allow both partial adjustment and a serially correlated disturbance term. It can be easily shown that the policy rule incorporating both partial adjustment and serial correlation can be expressed as in equation (4).

$$\begin{aligned} FR_t = {} & \alpha_0(1-s)(1-\rho) + (s+\rho)FR_{t-1} + (1-\rho) \\ & \{\alpha_\pi \overline{\pi}^c_{t,\overline{4}} + \alpha_y(\ln y_t - \ln y^*_t)\} - s\{(1-\rho)\alpha_\pi \overline{\pi}^c_{t-1,\overline{4}} \\ & +(1-\rho)\alpha_y(\ln y_{t-1} - \ln y^*_{t-1})\} - s\rho FR_{t-2} + \varepsilon_t. \end{aligned} \tag{4}$$

Note that if there is no serial correlation ($s = 0$ in [4]), we get the reduced-form policy rule shown in equation (2.3), and if there is no partial adjustment ($\rho = 0$ in [4]), we get the policy rule shown in (3.2). Of course, if both s and ρ are not zero, we have a policy rule with both partial adjustment and serial correlation.

In previous research, the forward-looking policy rule similar to the one given in equation (2.3) has often been estimated assuming rational expectations and using a generalized method of moments procedure (Clarida, Galí, and Gertler 2000). We follow this literature and estimate the policy rule assuming rational expectations; namely, we substitute actual future core inflation and actual current output gap for the expected inflation and output gap terms and use an instrumental variables procedure to estimate policy coefficients. However, we also estimate the policy rule using the Greenbook inflation forecasts as proxy for expected inflation. In contrast to previous work, we estimate the policy rule allowing for the presence of both interest-rate smoothing and serial correlation as in equation (4). We use a nonlinear instrumental variables procedure when rational expectations are assumed and nonlinear ordinary least squares procedure when the Greenbook forecasts are used. The instruments used are three lagged values of inflation, the federal funds rate, levels and first differences of the output gap, and the spread between the ten-year Treasury bond yield and the federal funds rate.

In previous work, as in Boivin (2006), ordinary least squares have been employed to estimate the Taylor rule that uses the Greenbook forecasts. However, the use of ordinary least squares requires the assumption that the Greenbook forecasts are contemporaneously uncorrelated with the policy shock ε_t. As noted in Boivin (2006), while some casual arguments can be made to support this assumption,[8] they cannot be directly verified, and hence would not be

[8] Reifschneider, Stockton, and Wilcox (1997) provide some information about the conditioning assumptions of the Greenbook forecasts over the last ten years. The first feature is that these forecasts are made under the typical assumption that the federal funds rate will remain unchanged

enough to convince a skeptic that the Greenbook forecasts may potentially be correlated with the policy surprise. This correlation may arise if the Greenbook forecasts reflect some contemporaneous information and the FOMC also reacts to such information by adjusting the policy rate, as argued in Rudebusch (2006). This endogeneity could introduce some bias in parameter estimates. In view of this consideration, we check the robustness of our results to the presence of potential endogeneity, using instrumental variables. In particular, we also estimate the Taylor rule, using the Greenbook forecasts made in previous quarters as instruments. We find our main results are robust with respect to this change in the estimation procedure.

Data

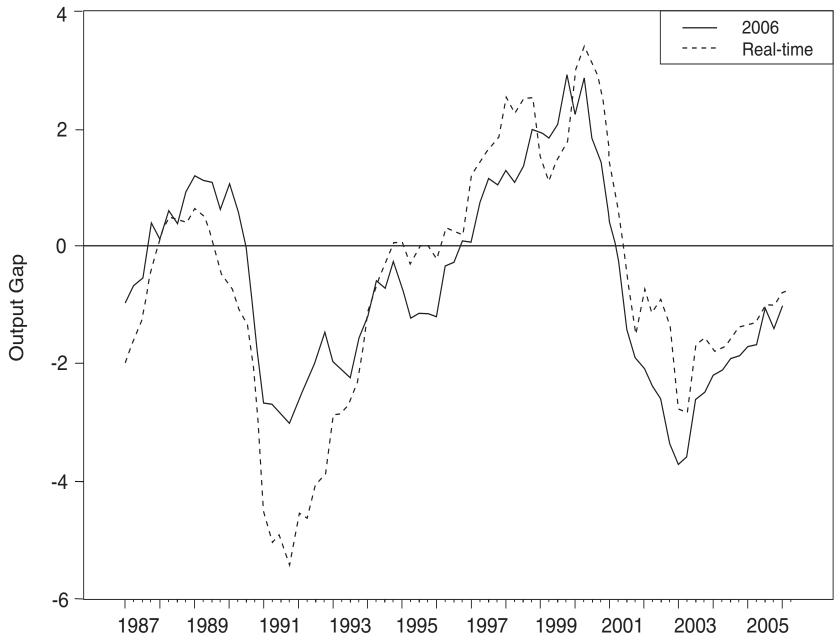
We estimate the policy rule in equation (4) over the period from 1987:1 to 2005:4 using the data on core CPI inflation and the output gap. For expected inflation, we also use the Greenbook inflation forecasts of core CPI inflation, prepared by the Board staff for the Federal Open Market Committee (FOMC) meeting held near the second month of the quarter. There is considerable evidence that the Greenbook forecasts are most appropriate in capturing policymakers' real-time assessment of future inflation developments. Romer and Romer (2000) show that the Federal Reserve has an informational advantage over the private sector, producing relatively more accurate forecasts of inflation than does the private sector. Bernanke and Boivin (2003) argue one needs a large set of conditional information to properly model monetary policy. In that respect, the Greenbook forecasts include real-time information from a wide range of sources, including the Board staff's "judgment," not otherwise directly measurable. The policy rule that uses the Greenbook forecasts is estimated over the period from 1988:1 to 2000:4.

Unlike inflation forecasts the Board staff's estimates of the output gap are not readily available. Here we follow the previous research using estimates of potential output prepared by the Congressional Budget Office (CBO).[9]

during the next six to eight quarters. This neutral assumption about the path of monetary policy may reflect the desire of the Board staff to avoid being construed as making policy recommendations, suggesting that for most of that period, the forecasts were not conditioned on the policy surprise. The second feature of these forecasts is a large "judgmental" component, making it hard for these forecasts to be mechanically reproduced by any particular forecasting model, thereby lessening the probability of a contemporaneous correlation between forecasts and the policy surprise.

[9] Potential output is defined as *trend in the productive capacity* of the economy and is estimated by the level of GDP attainable when the economy is operating at a high rate of resource use. The CBO estimates potential output for the economy, using a production function approach applied to each of five major sectors (nonfarm business, government, farm, household and nonprofit institutions, and residential housing) and then aggregating sectoral estimates of potential output. For example, for the nonfarm business sector CBO uses a neoclassical production function that relates output produced in that sector to labor (hours worked), capital, and total factor productivity. Potential output in nonfarm business sector is an estimate of output attainable when labor, capital,

Figure 1 Vintage 2006 and Real-Time Output Gap (Congressional Budget Office)



However, we also construct a real-time series on the output gap using the Congressional Budget estimates of actual and potential output series available in real time.[10] Unlike the data on the output gap, the data on CPI is not significantly revised, and hence we use the 2006 vintage dataset for core CPI.

Figure 1 charts real-time estimates of the output gap from 1987 to 2005. The most recent vintage (2006) estimates of the output gap are also charted. The main observation is that the real-time estimates of the output gap are not too different from their recent vintage estimates with the exception of periods 1990 to 1993 and 1995 to 1998. The real-time estimates of the output gap during the period surrounding the 1990–1991 recession indicate the presence of considerably more slack in the economy than what is indicated by current

and total factor productivity variables in the production function are set at their cyclically adjusted levels (Congressional Budget Office 2001).

[10] In January of each year from 1991 to 2006, the Congressional Budget Office has released the historical data on actual and potential output. For the period 1987 to 1990, the output gap is constructed using the series on actual and potential output given in the 1991 vintage data file. For 1991, we have used the pertinent series on actual and potential output from the 1992 vintage data file and for each year thereafter. So, the potential output estimate for 2005 is constructed using the data file released in January 2006.

Figure 2 Greenbook Forecasts and Actual Core CPI Inflation (Four-Quarter Average)

estimates. Hence, a policy rule that uses the real-time estimates of the output gap is likely to prescribe a lower funds rate target than what is indicated by the use of revised estimates. Similarly, real-time estimates of the output gap from 1995 to 1998 indicate far less slack in the economy than what is suggested by the current vintage estimates, due to the ongoing productivity acceleration that was not recognized by most economists at the time. Hence, for the subperiod 1995 to 1998 the funds rate target prescribed by the policy rule with the real-time output gap is higher than what is suggested by the current vintage estimate of the output gap, ceteris paribus. Given the size of output gap revisions, policy evaluation is likely to be affected whether one uses the real-time or revised data on the output gap.

Figure 2 charts the actual and Greenbook forecasts of the four-quarter-average core CPI inflation rate. As shown, the Greenbook forecasts track actual inflation fairly well, with the exception of periods, 1988:2 to 1989:2 and 1995 to 1997. In both these subperiods, the Greenbook was "too pessimistic" about future inflation. As some analysts have noted, during the first subperiod the Board staff may have worried about future inflation because the Greenspan Fed had kept interest rates low following the stock market crash of October

1987. During the second subperiod, productivity acceleration was underway, and most economists, including the Board staff, were slow in recognizing the favorable effects of productivity acceleration on inflation.

2. EMPIRICAL RESULTS

This section presents and discusses estimates of a Taylor rule fitted over the Greenspan period.

Estimates of Policy Response Coefficients

Table 1 presents estimates of policy response coefficients (α_π, α_y, ρ) from the Taylor rule in equation (4) estimated using the final as well as the real-time data on core CPI inflation and the output gap. Row 1 contains estimates derived using the current vintage data on the output gap, whereas row 2 contains estimates derived using the real-time data on the output gap. Row 3 contains ordinary least squares estimates using the Greenbook core CPI inflation forecasts and the real-time data on the output gap. We also present estimates of the first-order serial correlation coefficient s. The estimates in rows 1 through 3 of Table 1 suggest the following observations. First, all estimated policy response coefficients are correctly signed and statistically significant. In particular, the inflation response coefficient α_π is generally well above unity and the output response coefficient α_y is below unity, which suggests that the Greenspan Fed responded strongly to expected inflation and relatively weakly to output.

Second, the estimated serial correlation coefficient s is generally positive and statistically significant, indicating the presence of serially correlated errors in the estimated policy rules. As noted in Rudebusch (2006), the presence of serial correlation may reflect influences on the policy rate of economic variables to which the Federal Reserve may have responded but which are omitted from the estimated policy rule.

Third, even after allowing for the presence of serial correlation, the estimated partial adjustment coefficient ρ is positive and well above zero, which suggests the continued role of partial adjustment in generating a significant coefficient on the lagged value of the funds rate. This result is similar to that of English, Nelson, and Sack (2002). However, the magnitude of the estimated partial adjustment coefficient ρ reported here is somewhat smaller than what is found in previous research. As discussed later in this article, the point estimates of the partial adjustment coefficient range from .5 to .7 when the Taylor rule is alternatively estimated using the Greenbook forecasts of headline CPI and GDP inflation rates.

Table 1 Estimated Taylor Rules: Core CPI Inflation

Row	Period	Gap	Inflation	Estimation	α_π	α_y	ρ	s	$\overline{R}^2$	*SER*
1	2005:4	Revised	(Actual, FW)	IV	1.5	.78	.73	.59	.98	.326
					(3.7)	(3.2)	(6.7)	(4.3)		
2	2005:4	Real-time	(Actual, FW)	IV	2.2	.68	.66	.49	.98	.315
					(8.2)	(5.5)	(6.4)	(3.6)		
3	2000:4	Real-time	(GB, FW)	OLS	1.7	.64	.70	.35	.98	.257
					(8.8)	(6.4)	(13.1)	(2.5)		
4	2000:4	Real-time	(Actual, BW)	OLS	1.0	.73	.75	.59	.97	.329
					(1.8)	(2.6)	(5.9)	(3.5)		
5	2005:4	Real-time	(Actual, BW)	OLS	0.71	.51	0.66	.84	.98	.345
					(0.8)	(1.5)	(2.8)	(3.9)		

Notes: Rows 1 and 2 contain nonlinear instrumental variables (IV) estimates of policy coefficients from the forward-looking (FW) policy rule given below in (a) and use revised or real-time data on the output gap.

$$FR_t = \rho FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t,\overline{4}} + \alpha_y(\ ln\ \ y_t - ln\ \ y^*_t)\} + v_t. \tag{a}$$

Row 3 contains nonlinear ordinary least squares estimates (OLS) of policy coefficients from the forward-looking (FW) policy rule given below in (b) and use the Greenbook (GB) inflation forecasts of core CPI inflation and real-time CBO estimates of the output gap.

$$FR_t = \rho FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi GB\overline{\pi}^c_{t,\overline{4}} + \alpha_y(ln\ \ y_t - \ ln\ \ y^*_t)\} + v_t. \tag{b}$$

Row 4 contains nonlinear ordinary least squares (OLS) estimates of policy coefficients from the backward-looking (BW) policy rule given below in (c).

$$FR_t = \rho FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t-1} + \alpha_y(ln\ \ y_{t-1} - ln\ \ y^*_{t-1})\} + v_t. \tag{c}$$

The instruments used are three lagged values of the inflation rate, the funds rate, the output gap (final or real-time), the growth gap, and the spread between nominal yields on ten-year Treasury bonds and the federal funds rate. Parentheses contain t-values. *SER* is the standard error of estimate. Estimation was done allowing for the presence of first-order serial correlation in v_t, and s is the estimated first-order serial correlation coefficient. The sample periods begin in 1988:1 and end in the year shown in the column labeled "period."

These estimates indicate a faster convergence of the funds rate to its desired level over this sample period (see Panels A and B in Table 3).[11]

Forward- Versus Backward-Looking Specifications

The maintained hypothesis in this article is that the Greenspan Fed was forward looking, responding to expected inflation rather than lagged inflation. As noted at the outset, the original Taylor rule relates the actual federal funds rate to lagged inflation and the output gap. In order to investigate which specification better explains the Greenspan period, we also estimate the backward-looking specification. Rows 4 and 5 in Table 1 contain estimates of policy response coefficients from this backward-looking specification, using core CPI inflation and the real-time data on the output gap. Row 4 reports estimates for the subperiod 1988:1 to 2000:4, as does row 5 for the complete sample period 1988:1 to 2005:4.

One key feature of the backward-looking specifications reported in Table 1 is that the estimated inflation response coefficient α_π is close to or below unity and not always statistically significant. These estimates suggest that the Greenspan Fed did not respond strongly to inflation.[12] This conclusion is in sharp contrast to the one suggested by forward-looking specifications, according to which the Greenspan Fed responded strongly to inflation.

How does one decide which one of these two alternative specifications better describes the Greenspan period? The first to note is that the forward-looking specification better fits the data, because the forward-looking specification based on the Greenbook forecasts has a lower standard error of estimate than the backward-looking specification, (compare SERs across rows 3 and 4 in Table 1). We investigate this issue further by testing the validity of alternative specifications, using a general specification that nests both backward- and forward-looking specifications. In particular, consider a general specification

[11] As illustrated in Rudebusch (2006), the typical estimate of the partial adjustment coefficient ρ for this sample period is .8, suggesting that if in response to changed economic conditions the Federal Reserve wanted to raise the funds rate by one percentage point, it would raise it by about 20 basis points in the first three months and by about 60 basis points after one year. Focusing on the Taylor rule, which is estimated using Greebook forecasts and real-time data on the output gap, the mid-point of the estimated range of the partial adjustment coefficient is .6, suggesting the adjustment of the actual funds rate to its desired level will be complete well before a year. See also English, Nelson, and Sack (2002), in which the use of real-time data in a forward-looking policy rule yields an estimate of the partial adjustment coefficient that is also quite low.

[12] Blinder and Reis (2005) report a similar finding. For the period from 1987:3 to 2000:1, they estimate a Taylor rule that relates the funds rate target to current inflation and the unemployment gap. The inflation response coefficient estimated during that time is .57, leading them to conclude that the Greenspan Fed did not respond strongly to inflation.

given in equation (5.1).

$$
\begin{aligned}
FR_t^* &= a + \alpha_\pi GB\overline{\pi}^c_{t,\bar{j}} + \alpha_y(\ln y_t - \ln y_t^*) + \alpha_{\pi 2}\overline{\pi}^c_{t-1} && (5.1)\\
&\quad + \alpha_{y2}(\ln y_{t-1} - \ln y_{t-1}^*), \\
FR_t &= \rho FR_{t-1} + (1-\rho)FR_t^* + v_t, \text{ and} && (5.2)\\
v_t &= s v_{t-1} + \varepsilon_t \ ,
\end{aligned}
$$

where all variables are defined as before. Equation (5.1) relates the federal funds rate target to variables suggested by both the specifications. The key assumption underlying the general specification (5.1) is that lagged inflation and the output gap may directly influence the current federal funds rate target, in addition to influencing it indirectly through the Greenbook inflation forecast. The backward-looking specification allows for the direct influence of lagged inflation and the output gap on the current funds rate target. If α_π and α_y are zero in (5.1), we get the backward-looking specification, and if $\alpha_{\pi 2}$ and α_{y2} are zero, we get the forward-looking specification.

Table 2 contains nonlinear ordinary least squares estimates of policy response coefficients from the general policy rule (5) estimated over the period from 1988:1 to 2000:4. In addition to using the four-quarter-average Greenbook inflation forecast, we also report estimates using the one-quarter and two-quarter-average inflation forecasts. As shown, estimated coefficients on the Greenbook forecast α_π and the current output gap α_y are correctly signed and statistically significant, whereas estimated coefficients on lagged inflation $\alpha_{\pi 2}$ and lagged output gap α_{y2} are not. The p-value of the null hypothesis that $\alpha_{\pi 2}$ and α_{y2} are zero is .89 to .94, leading to the conclusion that the data favors the forward-looking specification.[13]

Robustness Issues: Core Versus Headline Inflation and Ordinary Least Squares Versus Instrumental Variables

Another key aspect of the maintained hypothesis is that the Greenspan Fed was focused on core rather than headline inflation. Furthermore, the analysis using the Greenbook forecasts used ordinary least squares to estimate the Taylor rule. We now investigate the robustness of our results to a few changes in the specification of the Taylor rule and the choice of the estimation procedure.

Table 3 presents the Taylor rule estimated using the Greenbook forecasts of three alternative measures of inflation: core CPI, headline CPI, and the GDP implicit deflator. The measure of real-time output gap used is from

[13] The results do not change if the general specification is estimated including current values of inflation and the output gap, instead of lagged values of inflation. That is, the estimated coefficient on expected inflation remains significant and that on current inflation is not.

Table 2 Estimates of Policy Response Coefficients From a General Policy Rule: Core CPI Inflation

Row	GB Forecasts	α_π	α_y	ρ	$\alpha_{\pi 2}$	α_{y2}	SER	$\overline{R}^2$	p-value
1	1-q average	1.49	.91	.75	.29	-.1	.284	.97	.89
		(1.9)	(1.9)	(8.4)	(.3)	(.2)			
2	2-q average	1.85	0.86	0.75	.16	-.1	.275	.97	.94
		(2.0)	(1.9)	(8.5)	(.1)	(.3)			
3	4-q average	1.99	0.71	0.72	0.33	-.1	.262	.97	.93
		(2.5)	(2.0)	(8.3)	(.3)	(.3)			

Notes: The coefficients reported are nonlinear least squares estimates of the policy rule given below in (a) and use the Greenbook forecasts (GB) and real-time data on the output gap.

$$FR_t^* = a + \alpha_\pi GB\overline{\pi}^c_{t,\bar{4}} + \alpha_y(\ln\ y_t - \ln\ y_t^*) + \alpha_{\pi 2}\overline{\pi}_{t-1} + \alpha_{y2}(\ln\ y_{t-1} - \ln\ y_{t-1}^*), \qquad (a.1)$$

$$FR_t = \rho FR_{t-1} + (1-\rho)FR_t^* + v_t, \qquad (a.2)$$

where all variables are defined as in Table 1. Parentheses below coefficients contain t-values. The p-value reported is for the test of the null hypothesis that $\alpha_{\pi 2}$ and α_{y2} are zero. The sample period is from 1988:1 to 2000:4. We do not report the estimated serial correlation coefficient, though the equations are estimated assuming the presence of serial correlation.

the Congressional Budget Office and remains the same across these three inflation specifications. Panel A presents ordinary least squares estimates and Panel B, instrumental variables estimates. For a comparison, Panel C reports the Taylor rule estimated using actual future inflation and the final data on the output gap. The estimates presented in Table 3 indicate three main observations. First, focusing on the Taylor rule with the Greenbook forecasts, the hypothesis—the Greenspan Fed responded strongly to expected inflation and relatively weakly to the output gap—is robust with respect to the use of headline inflation forecasts and the instrumental variables procedure. The estimated inflation response coefficient is well above unity and the output gap response coefficient is below unity for all three measures of inflation. The instrumental variables estimates of key policy response coefficients yield conclusions that are qualitatively similar to those based on ordinary least squares estimates (compare estimates across Panels A and B). These results suggest that the bias in ordinary least squares estimates, introduced as a result of the potential endogeneity of the Greenbook forecasts, may be very small.

Second, as expected, the fit of the estimated Taylor rule as measured by the standard error of regression (SER) is somewhat worse if instrumental variables are used. However, the Taylor rule estimated with the Greenbook forecasts always has a lower standard error of regression than the Taylor rule estimated using actual future inflation and the revised data on inflation and on the output gap (compare the SERs across Panels A, B, and C).

Table 3 Estimated Taylor Rules

Sample Period	Inflation	α_0	α_π	α_y	ρ	s	$\overline{R}^2$	*SER*
		Panel A: Greenbook Forecasts/Ordinary Least Squares						
1988:1–2000:4	Core CPI	0.12	1.7	.64	.69	.35	.98	.257
		(0.20)	(8.8)	(6.4)	(13.1)	(2.5)		
1988:1–2000:4	CPI	-0.80	2.1	.81	.74	.46	.98	.253
		(0.70)	(6.4)	(5.5)	(14.4)	(3.3)		
1988:1–2000:4	GDP	0.70	1.9	.66	.66	.45	.98	.252
		(1.20)	(8.5)	(6.5)	(10.9)	(3.3)		
		Panel B: Greenbook Forecasts/Instrumental Variables						
1988:1–2000:4	Core CPI	-0.20	1.8	.62	.60	.45	.98	.270
		(0.30)	(9.3)	(6.6)	(6.9)	(3.2)		
1988:1–2000:4	CPI	-1.50	2.3	.78	0.64	.60	.98	.273
		(1.20)	(6.0)	(5.5)	(7.8)	(4.3)		
1988:1–2000:4	GDP	0.29	2.1	.64	.51	.55	.97	.278
		(0.40)	(9.0)	(7.2)	(4.5)	(4.1)		
		Panel C: Actual Future Inflation/Instrumental Variables						
1988:1–2000:4	Core CPI	2.70	1.0	.85	.80	.60	.97	.314
		(1.30)	(1.5)	(1.9)	(5.6)	(3.2)		
1988:1–2000:4	CPI	1.80	1.3	.80	.78	.61	.98	.324
		(1.00)	(2.2)	(2.0)	(7.3)	(3.4)		
1988:1–2000:4	GDP	1.30	1.9	.67	.72	.63	.96	.332
		(0.80)	(2.9)	(1.8)	(4.3)	(2.7)		

Notes: Panels A, B, and C contain nonlinear estimates of policy coefficients from the policy rule given below in (a). Panels A and B use the Greenbook inflation forecasts and the CBO real-time estimates of the output gap. Panel C uses actual future inflation and the final revised data on the output gap.

$$FR_t = \rho FR_{t-1} + (1-\rho)\{\alpha_0 + \alpha_\pi \overline{\pi}^c_{t,\overline{4}} + \alpha_y(\ln\ y_t - \ln\ y_t^*)\} + v_t. \quad \text{(a)}$$

The instruments used are three lagged values of the pertinent inflation variable: the federal funds rate, the output gap (real-time or final), the growth gap, and the spread between nominal yields on ten-year Treasury bonds and the federal funds rate. See notes in Table 1.

Third, regarding core versus headline inflation, the results are mixed. When the Greenbook forecasts are used, instrumental variables estimates favor the core CPI, whereas ordinary least squares estimates favor the headline GDP inflation (compare the SERs across Panels A and B in Table 3). However, as reported in the next section, when we compare the relative accuracy of the within-sample dynamic forecasts of the funds rate generated by these different Taylor rules, the Taylor rule with core CPI inflation forecasts yields slightly more accurate forecasts of the funds rate than the Taylor rule with headline inflation forecasts, supporting the maintained hypothesis that the Greenspan Fed was focused on core inflation.[14]

Predicting the Actual Path of the Federal Funds Rate Using the Greenbook Inflation Forecasts and the Real-Time Output Gap

In order to evaluate how well the forward-looking inertial Taylor rule estimated here predicts actual policy actions, we focus on the policy rule estimated using Greenbook core CPI inflation forecasts and the real-time CBO estimates of the output gap from 1987:4 to 2000:4. For this exercise we focus on ordinary least squares estimates. We carry out this evaluation in two alternative ways. According to the inertial Taylor rule estimated here, expected inflation (approximated by Greenbook inflation forecasts) and the output gap are two major determinants of the federal funds rate target. In order to see how well the actual funds rate is predicted by these two economic fundamentals, we generate the within-sample dynamic predictions of the funds rate from 1987:4 to 2000:4, using the estimated policy rule shown in equation (6).

$$FR_t^p = \hat{\rho} FR_{t-1}^p + (1 - \hat{\rho})\{\hat{\alpha}_0 + \hat{\alpha}_\pi GB\overline{\pi}_{t,\overline{4}}^c + \hat{\alpha}_y(\ln y_t - \ln y_t^*)\}, \qquad (6)$$

where FR^p is the predicted funds rate and other variables are defined as before. The key feature of the prediction equation (6) is that in generating the current-quarter predicted value of the funds rate, we use last quarter's predicted, but not actual value of the federal funds rate, in addition to using current-period values of two other economic fundamentals.

Figure 3 charts the within-sample dynamic predictions of the funds rate.[15] Actual values of the funds rate and the prediction errors are also charted. Two observations need to be highlighted. First, the actual funds rate has generally moved in the direction suggested by these two economic fundamentals (see Panel A). Second, the estimated policy rule predicts very well the actual level of the funds rate. The mean absolute error is .29 percentage points and the root mean squared error is .40 percentage points. Despite this good fit, however, there are few periods when the actual funds rate is far away from the value prescribed by economic fundamentals. Significant deviations, at least twice the root mean squared error, occur in 1988 and 1995 (see Panel B, Figure 3).

[14] We did not consider the consumption expenditure deflator (PCE) in this comparison, because the Federal Reserve only recently started focusing on core PCE. In fact, the Greenbook started producing forecasts of core PCE beginning in 2000, suggesting the Greenspan Fed was focused on core CPI for most of the period covered.

[15] The predictions begin in 1987:4. For generating the prediction for 1987:4, we use the preceding quarter's actual funds rate. For later periods, the predicted values are generated using the preceding period's predicted value and the current period estimates of expected inflation and the output gap.

Figure 3 Predicting the Actual Funds Rate

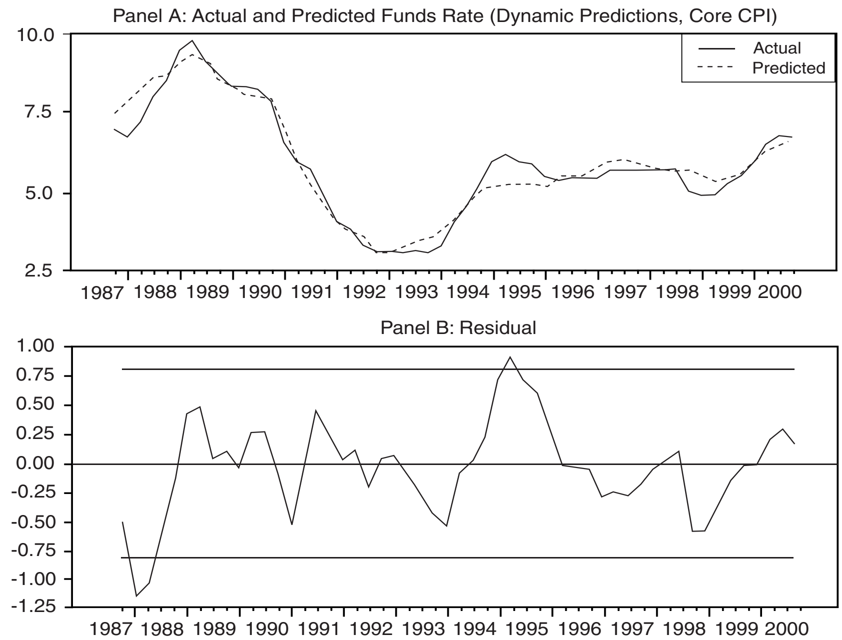


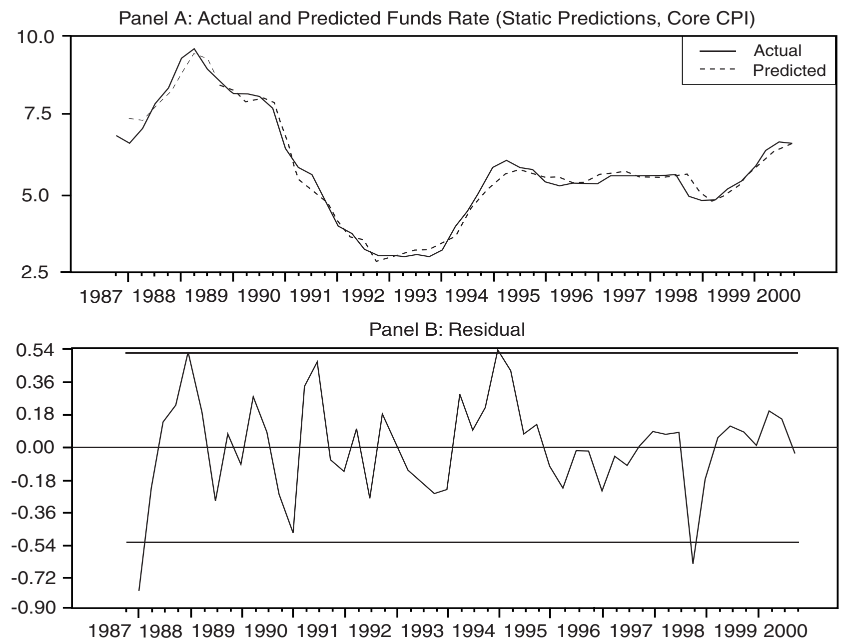
Figure 4 charts the static predictions of the federal funds rate, generated using the same policy rule but feeding in last quarter's actual value of the funds rate as shown below in equation (7).

$$FR_t^p = \hat{\rho} FR_{t-1} + (1 - \hat{\rho})\{\hat{\alpha}_0 + \hat{\alpha}_\pi GB\overline{\pi}^c_{t,\bar{4}} + \hat{\alpha}_y(\ln y_t - \ln y_t^*)\}. \quad (7)$$

In static forecasts the current-period forecast of the funds rate is determined, in part, by the current-period value of the desired policy rate suggested by economic fundamentals and, in part, by the one-period lagged value of the actual funds rate. So, in the static exercise the current forecast is influenced, in part, by actual policy actions, with the magnitude of the influence of policy on the forecast being determined by the size of the partial adjustment coefficient $\hat{\rho}$. Hence, the actual funds rate is likely better predicted by static than dynamic forecasts, because the latter are generated ignoring the recent history of actual funds rate changes.

A visual check of actual values of the funds rate and its static predictions charted in Figure 4 is consistent with the estimated policy rule. The mean absolute error is now .20 percentage points and the root mean squared error is

Figure 4 Predicting the Actual Funds Rate



.26 percentage points. Panel B charts the residuals. As shown, there are still a few periods of significant deviations. We see deviations at least as large as twice the root mean squared error occurring in 1988, 1989, 1995, and 1998:4. Thus, Figures 3 and 4 suggest that the Taylor rule estimated using Greenbook inflation forecasts and the real-time data on the output gap well predict actual policy actions, with the caveat that few episodes remain when the actual funds rate is significantly far from what is prescribed by this policy rule.

Using Actual Future Core Inflation and the Revised Output Gap

It is worth pointing out that in the prediction exercise the Taylor rule estimated using the Greenbook inflation forecasts and the real-time data on the output gap predicts actual policy actions better than the Taylor rule estimated using actual future inflation (core CPI) and the current vintage estimate of the output gap. In particular, we re-estimate the Taylor rule over the period from 1988:1 to 2000:4 and generate the within-sample, static and dynamic predictions of the funds rate, using the current vintage estimate of the output gap. For static predictions, the mean absolute error and root mean squared error are .30 and .37 percentage points, respectively. For dynamic predictions, the corresponding mean absolute error and the root mean squared errors are .72 and .84 percentage points. These prediction errors are substantially higher than those generated using the Greenbook inflation core CPI forecasts and the real-time output gap.

Core Versus Headline Inflation

The use of core inflation forecasts in the estimated Taylor rule produces slightly more accurate forecasts of the funds rate than those based on the headline inflation. For dynamic predictions of the funds rate generated using alternatively the Taylor rules based on core CPI, CPI, and GDP inflation forecasts, the mean absolute errors are .29, .35, and .33 percentage points, respectively. The corresponding root mean squared errors are .40, .44, and .41 percentage points. These summary statistics do favor core CPI, though the Taylor rule based on GDP inflation forecasts is a serious contender.[16,17]

Policy Residuals: Role of Additional Factors in the Estimated Taylor-Type Rule

As stated above, even though the use of Greenbook inflation forecasts and real-time data on the output gap enables the estimated policy rule to predict policy actions very well, there remain few periods when the actual funds rate is significantly away from values prescribed by the rule, with significant deviations occurring in 1988, 1989, 1995, and 1998:4. Many analysts contend that

[16] If the Taylor rules based on the Greenbook forecasts of three alternative measures of inflation—core CPI, CPI, and GDP—are estimated with instrumental variables, then the root mean squared errors generated by the dynamic prediction exercise are .46, .59, and .49 percentage points, respectively.

[17] It will be interesting to derive an estimate of the Greenspan Fed's inflation target under the additional assumption that the Fed's estimate of the short-term real rate can be approximated by the sample mean of the ex post real yield on three-month Treasury bills over a longer sample period, the latter defined as the nominal yield minus the lagged value of the four-quarter-average GDP inflation rate. By this metric, the short-term real rate is 1.9 percent if we use the sample period 1961:1–2005:4, and 2.1 percent if we use only the Greenspan period 1987:1–2005:4. These calculations suggest it is reasonable to assume that the Greenspan Fed's estimate of the short-real rate is approximately 2.0 percent. Given $rr^* = 2.0$ percent and given an estimate of the constant term from the estimated Taylor rule based on the Greenbook forecasts of core CPI inflation, the Greenspan Fed's inflation target calculated using the relationship $\hat{\alpha}_0 = rr^* + (1 - \hat{\alpha}_\pi)\pi^* \rightarrow .12 = 2.0 + (1 - 1.7)\pi^*$ is 2.7 percent. The result above—the Greenspan Fed's inflation target is 2.7 percent—may at first appear at odds with the 2.0 percent value assumed in the original Taylor rule, where inflation is measured by the behavior of GDP inflation. During the Greenspan era, GDP inflation has exhibited a somewhat different trend behavior than the core CPI inflation measure. Using the metric of comparing means, the sample mean of GDP inflation rates over 1987:1–2005:4 is 2.4 percent, which is lower compared with the value 3.0 percent computed using core CPI inflation over the same period. If we were to adjust the inflation targets for the presence of different means, then the Greenspan Fed having an inflation target of 2.7 percent based on the behavior of the core CPI inflation measure is equivalent to its having, instead, an inflation target of 2.1 percent based on the GDP inflation measure. The latter value is close to 2.0 percent assumed in the original Taylor rule.

significant deviations represent episodes when the Greenspan Fed responded to a variety of macroeconomic developments that are not included in the simple policy rule (Blinder and Reis 2005, Rudebusch 2006). To illustrate this point, consider the following narrative history of those developments.

The first episode occurs in 1988 and 1989. Following the stock market crash of October 1987, the Greenspan Fed kept interest rates low as an insurance against the heightened risk of a recession, so that in 1988 the actual funds rate is below what is prescribed by the Taylor-type rule. Inflation worries then may have led the Greenspan Fed to tighten more in 1989, which suggests that greater-than-policy-rule tightening in 1989 followed a somewhat looser policy of the previous year. Some support for this view emerges if we examine the Greenbook inflation forecasts in the period leading to 1989. As shown in Figure 2, for the period surrounding mid-to-late 1988 and early 1989, the Greenbook inflation forecasts turned out to be too pessimistic.

The second episode occurs in 1995 when the actual funds rate is higher than what is prescribed by the rule. The reasons for this greater-than-policy-rule tightening are not very clear. Taylor (2005) notes this may reflect preemptive policy tightening that began in 1994, whereas Rudebusch (2006) attributes it to an inflation scare that occurred at the end of 1994 evidenced by a rapid rise in long-term interest rates. Some limited support for the inflation scare argument appears in Figure 2, which shows that beginning in 1994:3, the Greebook inflation forecasts turned somewhat pessimistic about inflation.[18]

Finally, in 1998:4 the actual funds rate is below what is prescribed by the policy rule. This is the period when the international financial system was rocked by the Russian default and the demise of the Long-Term Capital Management (LTCM), which led the Greenspan Fed to lower interest rates. Together, these episodes suggest that the particular Taylor rule estimated in this article may not be considered a complete description of policy actions taken by the Greenspan Fed.

[18] Another factor that explains the greater-than-policy tightening in 1995 and in 1996–1997, as in some previous work that uses actual future inflation and the current vintage output gap measure, is the remarkable increase in productivity and potential output. At the time, most economists did not recognize these changes and, hence, may have overestimated the degree of utilization in product and labor markets, which likely reflected in tighter policy. However, a visual check of Figures 1 and 2 suggests that productivity acceleration may not be relevant in explaining the greater-than-policy tightening in 1995. As shown in Figure 1, real-time estimates of the output gap indicate far less slack in the economy than what is suggested by its 2006-vintage-only data in the subperiod following the year 1995. Similarly, the Greenbook forecasts become significantly pessimistic only in the years 1996–1997. Thus, these considerations suggest that while productivity acceleration may be relevant in explaining the post-1995 greater-than-policy tightenings documented in some previous work, its role in explaining the 1995 policy episode is in doubt.

3. CONCLUDING OBSERVATIONS

The main objective of this article is to investigate whether monetary policy actions taken by the Greenspan Fed can be summarized by a Taylor rule. Recent research highlights three aspects of the policy rule followed by the Greenspan Fed; namely, the Greenspan Fed was forward looking, focused on core inflation, and smoothed interest rates. The empirical work presented here supports the above-noted general characterization of the policy rule followed by the Greenspan Fed.

Using the Greenbook inflation forecasts and real-time Congressional Budget estimates of the output gap, this article reports evidence indicating that the Greenspan Fed reacted strongly to expected inflation and relatively weakly to the output gap. The evidence also indicates the Greenspan Fed smoothed interest rates, though the degree of interest-rate smoothing exhibited is considerably less than what is documented in previous research. The hypothesis that the Greenspan Fed was focused on core CPI inflation receives some support, as the Taylor rule based on the Greenbook forecasts of core CPI inflation does produce slightly more accurate forecasts of the funds rate than the Taylor rule that uses the Greenbook forecasts of headline CPI or GDP inflation.

This article finds that a Taylor rule estimated using the Greenbook core CPI inflation forecasts and real-time Congressional Budget estimates of the output gap predicts very well the actual path of the federal funds rate from 1987 to 2000. The Taylor rule estimated alternatively with the Greenbook GDP inflation forecasts seems to do as well. However, there are few periods when the Greenspan Fed is off the estimated rule, arising perhaps as a result of the Federal Reserve response to special macroeconomic developments not captured by the simple rule.

REFERENCES

Bernanke, Ben S., and Jean Boivin. 2003. "Monetary Policy in a Data-Rich Environment." *Journal of Monetary Economics* 50 (3): 525–46.

Blinder, Alan S., and Ricardo Reis. 2005. "Understanding the Greenspan Standard." Paper presented at the Federal Reserve Bank of Kansas City Economic Symposium, "The Greenspan Era: Lessons for the Future." Jackson Hole, WY.

Boivin, Jean. 2006. "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data." *Journal of Money, Credit and Banking* 38 (4): 1,149–173.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115 (1): 147–80.

Congressional Budget Office. 2001. "CBO's Method for Estimating Potential Output: An Update." The Congress of the United States (August).

English, William B., William R. Nelson, and Brian P. Sack. 2002. "Interpreting the Significance of the Lagged Interest Rate in Estimated Monetary Policy Rules." Mimeo, Federal Reserve Bank Board of Governors (April 24).

Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-time Data." *American Economic Review* 91 (4): 965–85.

Orphanides, Athanasios. 2002. "Monetary Policy Rules and the Great Inflation." *American Economic Review* 92 (2): 115–20.

Orphanides, Athanasios. 2006. "Monetary Policy Rules, Macroeconomic Stability and Inflation: A View from the Trenches." *Journal of Money, Credit and Banking* 36 (2): 151–75.

Reifschneider, David L., David J. Stockton, and David W. Wilcox. 1997. "Econometric Models and the Monetary Policy Process." *Carnegie-Rochester Conference Series on Public Policy* 47 (December): 1–37.

Romer, Christina, and David Romer. 2000. "Federal Reserve Information Advantage and the Behavior of Interest Rates." *American Economic Review* 90 (3): 429–57.

Rudebusch, Glenn D. 2002. "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia." *Journal of Monetary Economics* 49 (6): 1,161–87.

Rudebusch, Glenn D. 2006. "Monetary Policy Inertia: Fact or Fiction?" *International Journal of Central Banking* 2 (4): 85–135.

Taylor, John B. 1993. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39 (December): 195–214.

Taylor, John B. 2005. "Commentary: Understanding the Greenspan Standard." Remarks presented at the Federal Reserve Bank of Kansas City Economic Symposium, "The Greenspan Era: Lessons for the Future." Jackson Hole, WY.

Woodford, Michael. 2005. "Central Bank Communication and Policy Effectiveness." Paper presented at the Federal Reserve Bank of Kansas City Economic Symposium, "The Greenspan Era: Lessons for the Future." Jackson Hole, WY.

Economic Quarterly—Volume 93, Number 3—Summer 2007—Pages 251–286

Quantitative Models of Sovereign Default and the Threat of Financial Exclusion

Juan Carlos Hatchondo, Leonardo Martinez, and Horacio Sapriza

Business cycles in small emerging economies differ from those in developed economies. Emerging economies feature interest rates that are higher, more volatile, and countercyclical (interest rates are usually acyclical in developed economies). These economies also feature higher output volatility, higher volatility of consumption relative to income, and more countercyclical net exports.[1] Recent research is trying to develop a better understanding of these facts, as has been done for U.S. business cycles.

Because of the high volatility and countercyclicality of the interest rate, the (state-dependent) borrowing-interest rate menu is a key ingredient in any model designed to explain the cyclical behavior of quantities and prices in emerging economies. Some studies assume an exogenous interest rate.[2] Others provide microfoundations for the interest rate based on the risk of default.[3] This is the approach taken by recent quantitative models of sovereign default, which are based on the framework proposed by Eaton and Gersovitz (1981).[4] These articles build on the assumption that lenders can punish defaulting coun-

The authors would like to thank Kartik Athreya, Kevin Bryan, Huberto Ennis, and Ned Prescott for helpful comments. E-mails: JuanCarlos.Hatchondo@rich.frb.org; Leonardo.Martinez@rich.frb.org; and hsapriza@andromeda.rutgers.edu. The views expressed in this article do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] See Aguiar and Gopinath (2007), Neumeyer and Perri (2005), and Uribe and Yue (2006).

[2] See, for example, Aguiar and Gopinath (2007), Neumeyer and Perri (2005), Schmitt-Grohé and Uribe (2003), and Uribe and Yue (2006).

[3] Default episodes are not exceptional. Many nations have experienced episodes of sovereign default, some of the latest being Russia in 1998, Ecuador in 1999, and Argentina in 2001.

[4] See Aguiar and Gopinath (2006); Arellano (forthcoming); Arellano and Ramanarayanan (2006); Bai and Zhang (2006); Cuadra and Sapriza (2006a, b); Lizarazo (2005a,b); and Yue (2005).

tries by excluding them from international financial markets. The assumption is controversial on several grounds. First, it appears to be at odds with the existence of competitive international capital markets (which is assumed in these models). It is not obvious that competitive creditors would be able to coordinate cutting off credit to a country after a default episode.[5] Second, empirical studies suggest that once other variables are used as controls, market access is not significantly influenced by previous default decisions (see, for example, Gelos, Sahay, and Sandleris 2004, Eichengreen and Portes 2000, and Meyersson 2006).[6]

1. SUMMARY OF RESULTS

This article studies the role of the exclusion assumption for business cycle properties of emerging economies. It first describes the business cycle properties of a sovereign default model with exclusion and compares them with those of the same model without exclusion. The article finds that the presence of exclusion punishment is responsible for a high fraction of the sovereign debt that can be sustained in equilibrium. It also finds that the business cycle statistics of the model are not significantly affected by the exclusion punishment. The model without exclusion generates annual debt-output ratios of less than 2 percent. Whereas, the model with exclusion generates debt-output ratios between 4.8 and 6.3 percent. On the other hand, the cyclical behavior of consumption, output, interest rate, and net exports are not fundamentally different in the models with and without exclusion. An additional limitation shared by both model environments is that the volatility of the interest rate and (to a lesser extent) of the trade balance are too low compared to the data. This suggests that the exclusion assumption does not play an important role in these dimensions, and therefore future studies that do not rely on the threat of financial exclusion will not necessarily be handicapped in explaining the business cycle in emerging economies.

[5] This point is also raised by Cole, Dow, and English (1995) and Athreya and Janicki (2006).

[6] Sturzenegger and Zettelmeyer (2005) discuss how holders of defaulted bonds succeeded in interfering with cross-border payments to other creditors who had previously agreed to a debt restructuring. From this, they infer that holders of defaulted bonds may have been able to exclude defaulting economies from international capital markets. On the other hand, they conclude that "legal tactics are updated all the time, and new ways are discovered both to extract payment from a defaulting sovereign as well as to avoid attachments." In particular, they expect that "the threat of exclusion may be less relevant for some countries or to all countries in the future." For example, they explain that after Argentina defaulted in 2001, "attempts to actually attach assets have so far turned out to be fruitless." In any case, other forms of financing are always available to defaulting economies (issuing bonds at home, aid, official credit, multilateral or bilateral financing, etc.). The discussion in Sturzenegger and Zettelmeyer (2005) suggests, therefore, that defaulting economies might face at most a higher borrowing cost, though it is not clear how important this cost differential may be.

The model studied in this article builds on the framework studied in Aguiar and Gopinath (2006), which in turn, quantifies the model presented by Eaton and Gersovitz (1981). The most appealing feature about this setup is that it reduces the default decision to a simple tradeoff between current and future consumption without a major departure from the workhorse model used for real business cycle analysis in the last decades. Recent quantitative studies on sovereign default have shown that this environment can potentially account for important business cycle features in emerging economies and that it can be extended to address other issues (such as the optimal maturity structure of sovereign debt).[7] The framework studied in Aguiar and Gopinath (2006) is the simplest among the ones presented in recent studies. This has the advantage of making the discussion of the role of the exclusion assumption more transparent.[8] On the other hand, this has the disadvantage of hurting the performance of the model along several dimensions. Where appropriate, the article explains how the simplifying assumptions hurt the performance of the model.

This article studies a small open economy endowed with a single tradable good. As in Aguiar and Gopinath (2006), two endowment processes are considered: a process with shocks to the endowment *level* and a process with shocks to the endowment *growth rate*. The objective of the government is to maximize the present value of future utility flows of the representative agent. The government has only one financial instrument available: it can save or borrow using one-period bonds. These assets are priced in a competitive market inhabited by a large number of identical, infinitely lived, risk neutral-lenders. Lenders have perfect information regarding the economy's endowment. The government makes two decisions in every period. First, it decides whether to refuse to pay previously issued debt. Second, it decides how much to borrow or save. The baseline model features two costs of defaulting. First, the country may be excluded from capital markets. Second, it faces an "output loss." The endowment is reduced in a fixed percentage in the period following a default. The assumption that countries experience an output loss after a default intends to capture the disruptions in economic activity entailed by a default decision. IMF (2002), Kumhof (2004), and Kumhof and Tanner (2005) discuss how financial crises that lead to severe recessions are triggered by sovereign default.

---

[7] Arellano (forthcoming); Arellano and Ramanarayanan (2006); Bai and Zhang (2006); Cuadra and Sapriza (2006a,b); Lizarazo (2005a,b); and Yue (2005) extend the framework in Aguiar and Gopinath (2006) but maintain the basic assumptions (including the exclusion assumption).

[8] The only difference between the model presented in this article and the model in Aguiar and Gopinath (2006) is that here it is assumed that there is a unique period of output loss after default—in contrast with the stochastic number of periods of output loss assumed by Aguiar and Gopinath (2006). This allows us to eliminate the threat of exclusion without increasing the dimensionality of the state space. The Appendix shows that this departure does not have sizable effects on the results.

This article solves the model with and without the exclusion threat and compares their behavior. Mechanically, in the model with shocks to the endowment level, the default decision becomes relatively more sensitive to the endowment shock once the exclusion threat is eliminated and, therefore, less sensitive to the debt level. In turn, bond prices become less sensitive to the borrowing level. On the other hand, in the model with shocks to the growth rate the default decision becomes relatively less sensitive to the endowment shock, which increases the sensitivity of the bond price to the borrowing level. Given that in this class of models the high sensitivity of the default probability to the borrowing level limits their quantitative performance, the previous effects slightly improve the performance of the model with shocks to the endowment level and deteriorate the performance of the model with shocks to the growth rate. In spite of this, both models still do not replicate the default rates, the volatility of the trade balance, nor the volatility of the spread observed in the data.

The rest of the article proceeds as follows. Section 2 introduces the model. Section 3 presents the parameterization. Section 4 discusses the case in which the economy can be excluded from capital markets. Section 5 studies how the implications of the model change when the economy cannot be threatened with financial exclusion. Section 6 concludes the article.

2. THE MODEL

The environment studied in this article builds on the framework presented by Aguiar and Gopinath (2006), who study the quantitative performance of a model of sovereign default based on Eaton and Gersovitz (1981). Relative to Aguiar and Gopinath (2006), the only difference is that it is assumed here that there is a single period of output loss after default—in contrast with the stochastic number of periods of output loss assumed in their article. The Appendix shows that the results are not sensitive to this assumption.

The economy receives a stochastic endowment stream of a single tradable good. The endowment process has two components: a transitory shock and a trend shock, namely,

$$y_t = e^{z_t}\Gamma_t, \tag{1}$$

where y_t denotes the endowment realization in period t, z_t denotes the transitory shock, and Γ_t denotes the trend component.

The transitory shock z_t follows an AR(1) process with long-run mean μ_z, and autocorrelation coefficient $\left|\rho_z\right| < 1$, that is,

$$z_t = \left(1 - \rho_z\right)\mu_z + \rho_z z_{t-1} + \varepsilon_t^z, \tag{2}$$

where $\epsilon_t^z \sim N\left(0, \sigma_z^2\right)$.

The trend component evolves according to

$$\Gamma_t = g_t \Gamma_{t-1}, \tag{3}$$

where

$$ln\left(g_t\right) = \left(1-\rho_g\right)\left(ln\left(\mu_g\right) - m\right) + \rho_g ln\left(g_{t-1}\right) + \varepsilon_t, \tag{4}$$

$|\rho_g| < 1$, $\varepsilon_t \sim N\left(0, \sigma_g^2\right)$, and $\mu_g = \frac{1}{2}\frac{\sigma_g^2}{1-\rho_g^2}$.[9]

The objective of the government is to maximize the present value of future utility flows of the representative agent. The representative agent has preferences that display a constant coefficient of relative risk aversion:

$$u\left(c\right) = \frac{c^{(1-\sigma)} - 1}{1-\sigma},$$

where σ denotes the coefficient of relative risk aversion. Let β denote the discount factor. To ensure a well-defined problem it is assumed that

$$E\left\{\lim_{t\to\infty} \beta^t \left(y_t\right)^{(1-\sigma)}\right\} = 0.$$

The government makes two decisions in each period. First, it decides whether to refuse to pay previously issued debt. Second, it decides how much to borrow or save. As in previous quantitative studies, it is assumed that the government faces two penalties if it decides to default. One penalty is that it may be excluded from capital markets. The second penalty is that it faces an exogenous "output loss" of λ percent in the period following a default.

The exclusion state evolves as follows. In the default period, the economy is excluded from capital markets with probability $1-\phi_1$, with $\phi_1 \in [0, 1]$. In every period that follows a period of exclusion, the economy regains access to capital markets with probability $\phi \in [0, 1]$ or remains excluded for one more period with probability $1-\phi$.[10] This implies that the expected length of exclusion is given by $\frac{1-\phi_1}{\phi}$. If the economy was not excluded from financial markets at the end of the previous period, it is not excluded at the beginning of the current period.

The government can choose to save or borrow using one-period bonds. There is a continuum of risk-neutral lenders with "deep pockets." Each lender can borrow or lend at the risk-free rate r. Lenders have perfect information

---

[9] The endowment process is motivated by the work of Aguiar and Gopinath (2007). They find that shocks to trend growth (rather than transitory fluctuations around a stable trend) are the primary source of fluctuations in emerging markets.

[10] Previous quantitative studies of sovereign default assume that the government cannot borrow in the period it defaults ($\phi_1 = 0$), and it regains access to capital markets with a constant probability (ϕ) after that. In order to accommodate this possibility, it is assumed that ϕ_1 can be different from ϕ.

regarding the economy's endowment. The bond price is determined as follows. First, the government announces how many bonds it wants to issue. Then, lenders offer a price for these bonds. Finally, the government sells the bonds to one of the lenders who offered the highest price.

Let b denote the current position in bonds. A negative value of b denotes that the government was an issuer of bonds in the previous period. Each bond delivers one unit of good next period for a price of $q_d\left(b', z, \Gamma, g\right)$ this period. The price depends on the current default decision, d. This is due to the fact that a current default decreases future output and affects future default decisions.

The government compares two continuation values in order to decide whether to default or pay back the previously issued debt. The present discounted utility after a default is represented by $V_1(z, \Gamma, g, h)$. The variable h denotes the credit history of the government. It takes a value of 1 when the government defaulted in the previous period, and it takes a value of 0 when the government did not default in the previous period. The present discounted utility when all previously issued debt is paid back is represented by $V_0(b, z, \Gamma, g, h)$. The government defaults if the continuation value $V_1(z, \Gamma, g, h)$ is larger than $V_0(b, z, \Gamma, g, h)$ and does not default otherwise.

Let x denote the exclusion state. The variable x takes a value of 1 when the economy is excluded, and takes a value of 0 otherwise. Let $V(b, z, \Gamma, g, h, x)$ denote the government's value function at the beginning of a period.

In the period it defaults, the economy can or cannot be excluded from financial markets. Let $\tilde{V}_1(z, \Gamma, g, h, 0)$ denote the continuation value when the economy defaulted and is not excluded. Let $\tilde{V}_1(z, \Gamma, g, h, 1)$ denote the continuation value when it is excluded. Thus,

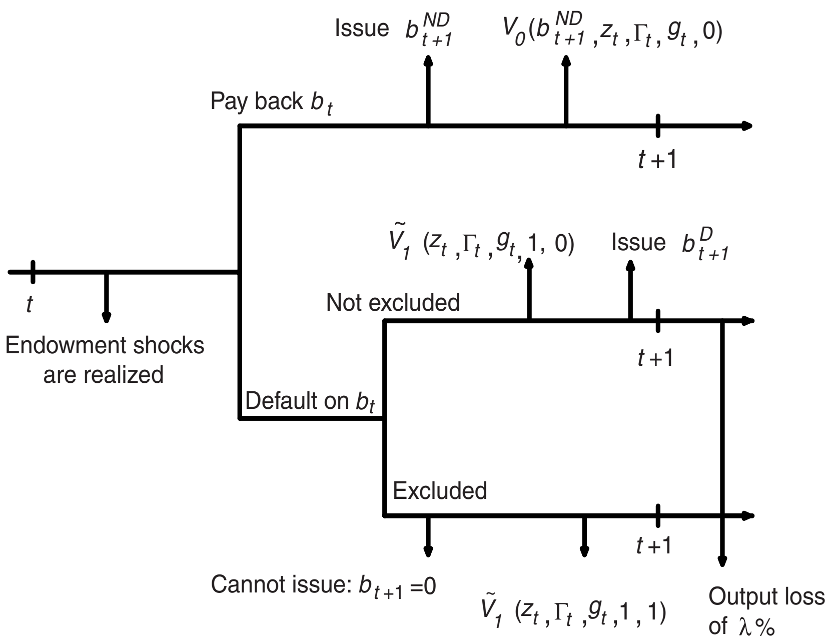
$$V_1(z, \Gamma, g, h) = \phi_1 \tilde{V}_1(z, \Gamma, g, h, 0) + (1 - \phi_1)\, \tilde{V}_1(z, \Gamma, g, h, 1). \qquad (5)$$

The timing of the decisions within a period is summarized in Figure 1. At the beginning of the period the endowment shocks are realized. The realization in period t of a state variable x is denoted by x_t. After observing the endowment realization, the government decides whether to pay back previously issued debt. If it decides to pay the debt back, the government issues an amount b_{t+1}^{ND} of bonds and faces a continuation value of $V_0\left(b_{t+1}^{ND}, z_t, \Gamma_t, g_t, 0\right)$. If the government defaults, it may or may not be excluded from capital markets today. If it is not excluded, it faces a continuation value of $\tilde{V}_1(z_t, \Gamma_t, g_t, 1, 0,)$. If it is excluded, it faces a continuation value of $\tilde{V}_1(z_t, \Gamma_t, g_t, 1, 1,)$. If the government defaults and is not excluded today from capital markets, it issues an amount b_{t+1}^{D} of bonds. After a default, the government faces an output loss of λ percent in period $t+1$ regardless of whether it was excluded from capital markets in period t.

The value function of a defaulting economy that is excluded in the default period is computed as follows:

$$\tilde{V}_1(z, \Gamma, g, h, 1) = u\left(y\left(1 - h\lambda\right)\right) + \beta E\left[V(0, z', g'\Gamma, g', 1, e)\right], \qquad (6)$$

Figure 1 Order of Events and Alternative Continuation Values in Period t



where

$$E\left[V(0,z',g'\Gamma,g',1,e)\right] = \int\int \left[\begin{array}{c} \phi V(0, z', g'\Gamma, g', 1, 0)+ \\ (1-\phi)\, V(0, z', g'\Gamma, g', 1, 1) \end{array} \right] F_z\left(dz' \mid z\right) F_g\left(dg' \mid g\right).$$

If the government has decided to default and is excluded in the period of default, it consumes the aggregate endowment (there are no financial transfers from or to the rest of the world) and carries zero debt to the next period. At the beginning of the following period, exclusion finishes with probability ϕ. The expected continuation value of this scenario is $V(0, z, g\Gamma, g, 1, 0)$. If the exclusion time is extended, the expected continuation value is $V(0, z, g\Gamma, g, 1, 1)$.

The dynamic programming problem for a defaulting economy that is not excluded in the default period is

$$\tilde{V}_1(z, \Gamma, g, h, 0) = \max_{b'} \left\{ \begin{array}{c} u\left(y(1-h\lambda) - q_1\left(b', z, \Gamma, g\right) b'\right) + \\ \beta E\left[V(b', z', g'\Gamma, g', 1, 0)\right] \end{array} \right\} \tag{7}$$

where

$$E\left[V(b', z', g'\Gamma, g', 1, 0)\right] = \int\int V(b', z', g'\Gamma, g', 1, 0) F_z\left(dz' \mid z\right) F_g\left(dg' \mid g\right).$$

In this case, the government must choose how much debt it will issue.

The value function of the government when it has decided to pay back its debt is obtained from the following Bellman equation:

$$V_0(b, z, \Gamma, g, h) = \max_{b'} \left\{ \begin{array}{c} u\left(y(1-h\lambda) + b - q_0\left(b', z, \Gamma, g\right) b'\right) + \\ \beta E\left[V(b', z', g'\Gamma, g', 0, 0)\right] \end{array} \right\}, \quad (8)$$

where

$$E\left[V(b', z', g'\Gamma, g', 0, 0)\right] = \int\int V(b', z', g'\Gamma, g', 0, 0) F_z\left(dz' \mid z\right) F_g\left(dg' \mid g\right).$$

The function $V(b, z, \Gamma, g, h, x)$ is computed as follows:

$$V(b, z, \Gamma, g, h, 0) = \max\{V_1(z, \Gamma, g, h), V_0(b, z, \Gamma, g, h)\}, \quad (9)$$

and

$$V(b, z, \Gamma, g, h, 1) = u(y(1-h\lambda)) + \beta E\left[\phi V(0, z', g'\Gamma, g', 0, e)\right],$$

where

$$E\left[V(0,z',g'\Gamma,g',0,e)\right] = \int\int \left[\begin{array}{c} \phi V(0, z', g'\Gamma, g', 0, 0) + \\ (1-\phi) V(0, z', g'\Gamma, g', 0, 1) \end{array} \right] F_z\left(dz' \mid z\right) F_g\left(dg' \mid g\right).$$

Let

$$d(b, z, \Gamma, g, h) = \left\{ \begin{array}{ll} 1 & \text{if } V_1(z, \Gamma, g, h) > V_0(b, z, \Gamma, g, h) \\ 0 & \text{if } V_1(z, \Gamma, g, h) \le V_0(b, z, \Gamma, g, h) \end{array} \right. \quad (10)$$

denote the equilibrium default decision.

The price of a bond if a default decision d was made in the current period satisfies the lenders' zero profit condition. It is given by

$$q_d\left(b', z, \Gamma, g\right) = \frac{1}{1+r}\left[1 - E\left[d' \mid b', z, \Gamma, g, d\right]\right], \quad (11)$$

where

$$E\left[d' \mid b', z, \Gamma, g, d\right] = \int\int d\left(b', z', g'\Gamma, g', d\right) F_z\left(dz' \mid z\right) F_g\left(dg' \mid g\right)$$

denotes the probability that the government decides to default if it purchases b' bonds, and the current default decision is d.

Equilibrium Concept

Definition 1 *A recursive competitive equilibrium is characterized by*

1. a set of value functions $V(b, z, \Gamma, g, h, x)$, $V_1(z, \Gamma, g, h)$, and $V_0(b, z, \Gamma, g, h)$;

2. a set of policies for asset holdings $b'_0\,(b, z, \Gamma, g, h)$ and $b'_1\,(b, z, \Gamma, g, h)$, and a default decision $d\,(b, z, \Gamma, g, h)$; and

3. a bond price function $q_d\left(b', z, \Gamma, g\right)$,

such that

(a) $V\,(b, z, \Gamma, g, h, x)$, $V_1\,(z, \Gamma, g, h)$, and $V_0\,(b, z, \Gamma, g, h)$ satisfy the system of functional equations (5)–(9);

(b) the default policy $d\,(b, z, \Gamma, g, h)$ and the policies for asset holdings $b'_0\,(b, z, \Gamma, g, h)$ and $b'_1\,(b, z, \Gamma, g, h)$ solve the dynamic programming problem specified by equations (5)–(9); and

(c) the bond price function $q_d\left(b', z, \Gamma, g, h\right)$ is given by equation (11).

Discussion of the Environment

The model analyzed in this article relies on several simplifying assumptions. This has the advantage that the model remains tractable and that the main mechanisms can be presented in a more transparent way. The disadvantage of using such a stylized framework is that the model is ill-suited to account for the quantitative behavior of some key variables.[11] The rest of this section discusses several simplifications embedded in the environment presented in the previous section and extensions that have been studied in the literature.

Focusing on an endowment economy simplifies the analysis. A more comprehensive study of the business cycle would require incorporating capital and labor into the model. Aguiar and Gopinath (2006) also consider an extension of the basic model with labor as the only input in the production function. The results do not change significantly. More recently, Bai and Zhang (2006) study a production economy with capital.

The model assumes that the government issues one-period bonds. Allowing the government to issue long-term bonds would introduce nontrivial complications to the analysis. For instance, if the government can issue two-period bonds, it is necessary to keep track of how much debt was issued two periods ago (which is due today) and how much debt was issued one period ago (which will be due tomorrow). Alternatively, if the government only issues annuities there would only be one state: how many annuities have been issued since the last default. However, the pricing of the annuities issued today would be more complex than the price of a one-period bond. Lenders would not only need to compute the probability of a default in the following period, but also the probability of a default two-periods ahead, conditional on not observing a default tomorrow; the probability of observing a default three-periods ahead;

[11] Other authors have studied different extensions of this framework, which have improved its quantitative performance. See, for example, Arellano (forthcoming); Arellano and Ramanarayanan (2006); Bai and Zhang (2006); Cuadra and Sapriza (2006a, b); Lizarazo (2005a, b); and Yue (2005).

conditional on not observing a default in the next two periods; and so on. Arellano and Ramanarayanan (2006) allow the government to issue short and long bonds.

It was assumed that the government cannot issue bonds contingent on the future realization of its endowment. Even if creditors have perfect information regarding the economy's endowment, this would not imply that contracts contingent on the endowment realization could be written (the endowment may not be verifiable). In reality, one limitation for writing contracts contingent on real variables is that the government could manipulate the measurement of these variables (see Borensztein and Mauro 2004). Determining to what extent bonds can be state contingent in reality and studying some degree of state contingency in quantitative models of sovereign default are interesting avenues for future research.

The assumption that countries experience an exogenous output loss after defaulting intends to capture the disruptions in economic activity entailed by a default decision. In general, default episodes are not observed in economic booms but in recessions. This means that a fraction of the low economic activity that is observed after a default episode can be explained by weak fundamentals pre-existing the default decision. Thus, not all of the decrease in economic activity observed after a default is related to the default decision and cannot be considered as a cost of defaulting. On the other hand, default decisions are likely to introduce disruptions in economic activity of the defaulting economy. IMF (2002), Kumhof (2004), and Kumhof and Tanner (2005) discuss how financial crises that lead to severe recessions are triggered by sovereign default. This is due to the fact that government debt is not only held by foreigners but also by locals, and in particular by local banks—something that is not explicitly considered in the stylized model studied in this article. Thus, government default may hurt financial intermediation significantly (see IMF 2002 for a discussion of recent episodes).[12]

In the model, the output loss triggered by the default decision is independent of the size of the default. If the output loss represents the damage made by the default decision through the local financial system, it could be argued that the loss should depend (positively) on the amount that is not repaid by the government (in particular, it should depend on the amount held by the locals; see, for example, IMF 2002). Considering this would introduce additional complications to the analysis though it is an interesting avenue to be pursued

[12] In the stylized model discussed in this article, the output loss of λ percent in the period after a default intends to capture the cost of defaulting implied by the disruptions in economic activity triggered by the declaration of default. The calibration of the parameter λ should capture these disruptions and does not intend to match the overall decrease in output observed after a default. In contrast to the exclusion from capital markets, the output loss does not intend to capture a punishment imposed by creditors. Eaton and Gersovitz (1981) discuss output loss as the result of punishments.

in future work. If the output loss depends on the amount not paid by the government, it can be argued that the government should also be allowed to choose the size of the default.[13] Arellano (forthcoming) argues that the output loss depends on the state of the economy and, thus, introduces this into the model.

Previous quantitative studies assume that after default, the economy suffers the output loss for a stochastic number of periods (the periods in which the economy is excluded from capital markets). For simplicity, this assumption is modified in this article. Assuming that the output loss lasts for a stochastic number of periods in a context in which there is no financial exclusion raises the possibility of scenarios in which the government defaults before the duration of output losses triggered by the previous default has ended. This would require keeping track of the number of output losses the economy is suffering and would increase the dimensionality of the state space. The Appendix shows that the results are not sensitive to the modification of the output loss process utilized in this article.

The assumption that countries are excluded from capital markets after a default episode is motivated by evidence of a drainage in capital flows into countries that defaulted (see, for example, Gelos, Sahay, and Sandleris 2004). However, it very well may be that the difficulties in market access observed after a default episode respond to the same factors that triggered the default decision itself.[14] In support of this, Gelos, Sahay, and Sandleris (2004) document that once other variables are used as controls, market access is not significantly influenced by previous defaults (see also Eichengreen and Portes 2000 and Meyersson 2006). Moreover, it is not obvious that after a default episode competitive creditors would be able to coordinate cutting off credit to defaulting countries. Thus, the study of an environment in which a defaulting economy cannot be excluded from capital markets is the first building block of any work that attempts to explain the exclusion outcome as an endogenous outcome of the model.

[13] In this article, the government must decide whether it honors all the debt issued in the previous period or whether it defaults on all of it. But this is not a restrictive assumption given that the costs of defaulting are orthogonal to the amount of debt that is repudiated. In this case, the government would never find it optimal to default on less than a 100 percent of the outstanding debt. This would not be the case if the cost of defaulting depends on the amount repudiated. Yue (2005) studies partial default in an environment in which the defaulted amount is decided in a bargaining process between the government and the lenders.

[14] For example, Hatchondo, Martinez, and Sapriza (2006b) analyze a model in which both default and the difficulties in market access after default may be triggered by a change in the policymaker in power.

Table 1 Parameter Values Specific to Models I and II

	Model I	Model II
ρ_g	–	0.17
σ_g	0%	3%
ρ_z	0.90	–
σ_z	3.4%	0%

Notes: A period in the model corresponds to a quarter.

3. PARAMETERIZATION

The model is solved numerically using value function iteration and interpolation as in Hatchondo, Martinez, and Sapriza (2006a).[15] Whenever possible, this article considers the same parameter values as in Aguiar and Gopinath (2006), which facilitates the comparison of the results. To solve the model numerically, Bellman equations are first recast in detrended form. All variables are normalized by $\mu_g \Gamma_{t-1}$ as in Aguiar and Gopinath (2006). This normalization implies that the mean of the detrended endowment is one.

Even though Aguiar and Gopinath (2007) argue that the best representation of the output process for emerging economies is characterized by equations (1) through (4), this specification requires keeping track of z and g as state variables. The computational method used in previous articles in the literature does not allow solving for this specification without incurring sizable approximation errors. Instead, they consider two alternative endowment processes. In Model I, the economy is hit only with transitory shocks (z shocks). In Model II, the economy is hit with shocks to the trend only (g shocks). Table 1 reports the parameter values specific to each of the two model alternatives.

Parameters values that are common across models are presented in Table 2. Aguiar and Gopinath (2006) assume an output loss of 2 percent during the exclusion period. This is based on empirical estimates of the output loss triggered by a default decision (see Chuhan and Sturzenegger 2005). As explained above, for simplicity this article assumes that all output loss occurs only in one period, the period that follows the decision. The value of λ is calibrated to make the output-loss cost of defaulting in this article equivalent to the one in Aguiar and Gopinath (2006). In particular, the value of λ is chosen to be such that for Model I, the mean debt level in the simulations is the same as the one in the original formulation of Aguiar and Gopinath (2006). We show that this value (of λ) enables Model II to generate a similar level of debt as in Aguiar and Gopinath (2006).

[15] The value functions V_0 and V_1 are approximated using Chebychev polynomials. Fifteen polynomials on the asset space and ten on the endowment shock are used. Results are robust to using more polynomials.

Table 2 Parameter Values Common to Models I and II

Risk Aversion	σ	2
International Interest Rate	r	1%
Probability of Redemption in the Same Period of Default	ϕ_1	0%
Probability of Redemption	ϕ	10%
Mean Growth Rate	μ_g	1.006
Mean (log) Transitory Productivity	μ_z	$-0.5\sigma_z^2$
Discount Factor	β	0.8
Loss of Output	λ	8.3%

Notes: A period in the model corresponds to a quarter.

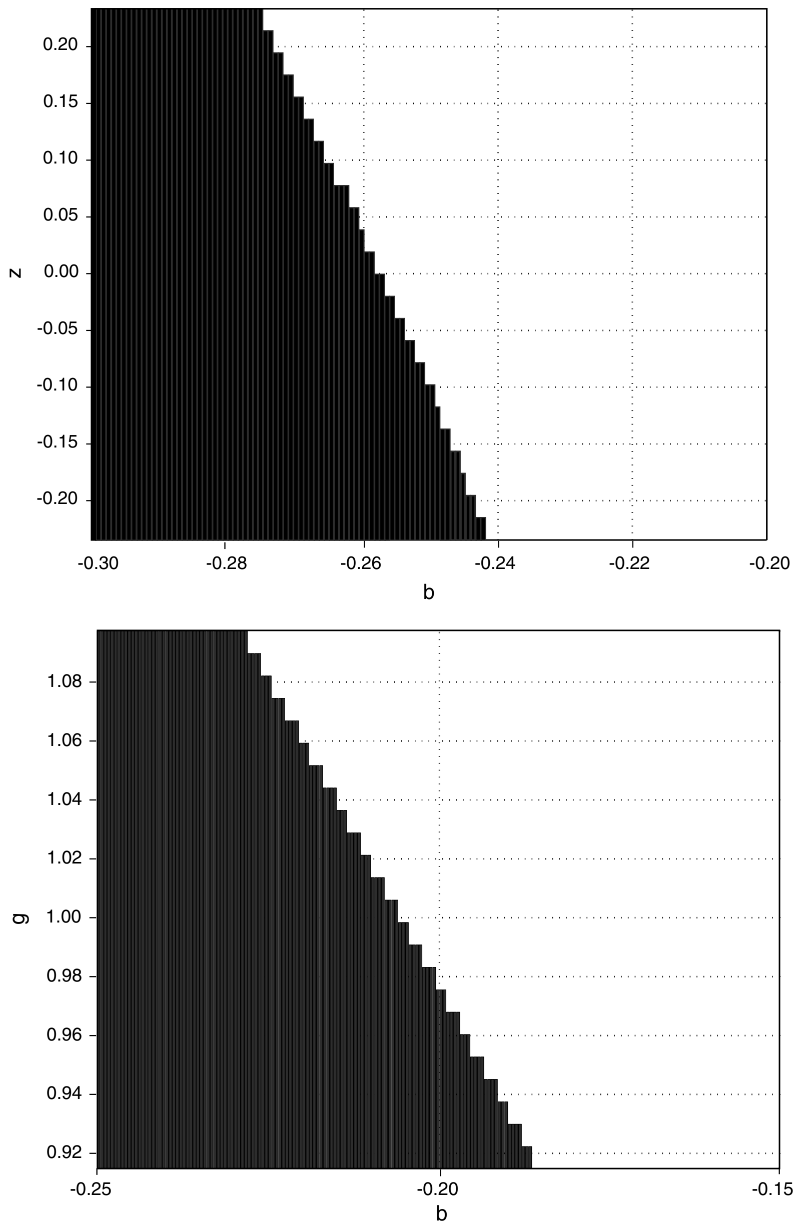
Except for the value of λ, the remaining parameters take the same values as in Aguiar and Gopinath (2006). The coefficient of relative risk aversion of 2 is within the range of accepted values. The probability of redemption implies an average autarky duration of 2.5 years (in the model, a period refers to a quarter), similar to the value estimated by Gelos, Sahay, and Sandleris (2004)—Section 5 presents the results when creditors cannot exclude a defaulting economy. The process of output is calibrated to match the process for Argentina from 1983 to 2000. The subjective discount factor is set to 0.8. This departs from standard macro models. As in Aguiar and Gopinath (2006), in the stylized framework discussed in this article, a low discount factor is needed to induce the economy to accumulate debt and be willing to accept a higher spread over the risk-free interest rate (the international interest rate). The limitations faced by the stylized framework to generate default with more reasonable discount factors may be a consequence of its simplifying assumptions.[16] As was mentioned previously, recent articles study different extensions of this framework that improve its quantitative performance (assuming that lenders can use financial exclusion as a punishment).

4. RESULTS WITH EXCLUSION

This section presents the results obtained when the parameters that determine the exclusion process are chosen as in Aguiar and Gopinath (2006), that is, $\phi_1 = 0$ and $\phi = 0.1$. This means that the government cannot issue bonds in the period it defaults, and after that it faces a constant probability ϕ of regaining access to capital markets.

[16] For example, the end of Section 4 describes how the assumption that the government can only issue one-period bonds increases the marginal issuance cost. If it were not for the low discount factor, the issuance volume and spreads observed in equilibrium would be even lower than what is observed in the data.

Figure 2 Default Regions



Notes: Default region when the endowment is hit with transitory shocks (top panel) and growth trend shocks (bottom panel).

Equilibrium Default Region and Bond Prices with Exclusion

The shaded areas in Figure 2 display the default regions (i.e., the combinations of endowment shocks and debt levels for which the economy would choose to default) of the model with transitory and trend shocks.[17] Both graphs show that the higher the endowment shock, the higher the minimum debt level at which it is optimal to declare a default. From another perspective, for a given initial debt level, the government finds it optimal to default only if the endowment shock is sufficiently low.

The benefit of defaulting is that resources that would have been allocated to pay back previously issued bonds are, instead, allocated to current consumption. There are two costs entailed by a default decision: a loss in output and the inability of the government to use international capital markets to smooth out domestic endowment shocks. It should be noticed that the "costs" of defaulting do not depend on the debt level at the time of default. Thus, a higher initial debt level increases the benefits of a default without increasing the costs. For sufficiently large debt levels, the benefits of defaulting offset the "fixed" costs. This explains why in Figure 2 it is optimal for the government to default on relatively large values of debt (low b).

A low endowment shock implies that there are less resources available in the current and subsequent periods.[18] Given that the output loss that follows a default decision is a constant fraction of the underlying potential output, it is more costly to default for high endowment realizations than for low endowment realizations. In the model with transitory shocks, this is the main force behind the negative relationship between the shock to the endowment and the debt threshold at which the government is indifferent between defaulting and not defaulting.[19] However, in the model with trend shocks, a high shock today signals higher growth rates in the future. This increases the desire to borrow as it allows bringing future resources to the current period. The fact that the ability to borrow is more valuable in good than in bad times helps explain why the government defaults only on larger debt volumes in good times.

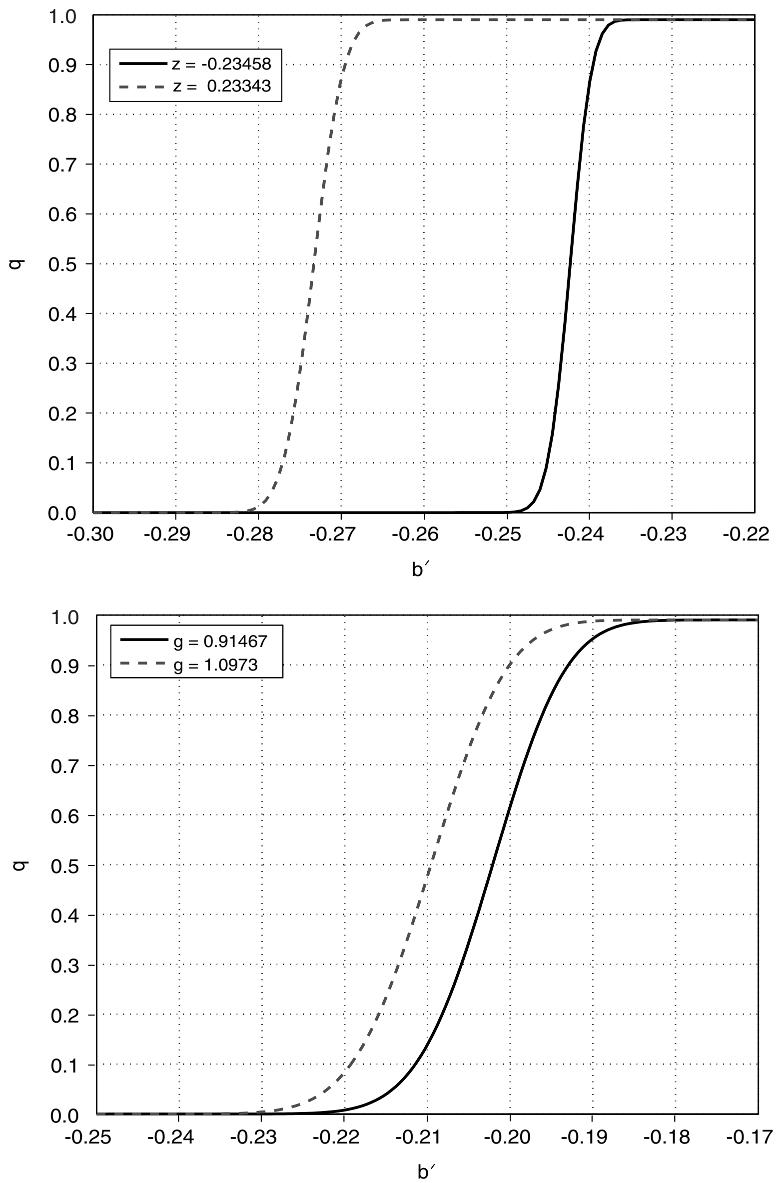
Figure 3 shows the equilibrium bond prices faced by the government as a function of the current issuance level ($-b$) and the endowment shock. The

[17] The detrended output process has a mean of 1. This implies that the debt levels b in Figure 2 correspond to the ratio of debt-to-mean output.

[18] Both the trend shocks and the transitory shocks display positive autocorrelation.

[19] If the endowment is low, the marginal utility of consumption is high, and therefore the gain from defaulting is high. However, if the model with transitory shocks is solved assuming that the output loss is a fixed amount instead of a percentage of output, the default region becomes almost vertical but with a positive slope. This is due to the fact that with a high endowment shock the economy displays a less intense desire to issue debt, and therefore it assigns a lower value to retaining access to capital markets.

Figure 3 Bond Price Menus



Notes: Equilibrium bond price menu faced by the government at low and high endowment realizations. The low endowment realization is three standard deviations to the left of the unconditional mean. The high endowment realization is three standard deviations to the right of the unconditional mean. The top panel shows the equilibrium bond price in the model with shocks to the level. The bottom panel shows the equilibrium bond price in the model with shocks to the trend.

curves have a waterfall shape. For relatively low issuance levels there is no risk of default. In this case, competitive investors demand the risk-free rate in compensation for purchasing the government's bonds. For issuance volumes for which there is a positive probability of default tomorrow, the rate of return demanded for holding bonds is higher than the risk-free rate, i.e., the price offered is lower than $\frac{1}{1+r}$. Finally, for sufficiently high issuance volumes, it is common knowledge that the government would default in the following period for almost any endowment realization. In this case, investors offer a zero price for each bond issued today.

It should be noted that price q is nondecreasing in the current endowment realization. In other words, the higher the endowment, the higher the issuance level at which the price starts to fall. This is due to the persistence in the endowment process and the shape of the default regions. A higher endowment today implies that it is more likely to observe high endowments in the following period, and therefore it makes the default probability lower.

Business Cycle Properties With the Exclusion Punishment

The model is simulated for 750,000 periods (500 samples of 1,500 observations each). In order to compute business cycle statistics, 400 samples of the last 72 periods before a default episode are used. The samples selected are such that the last exclusion period was observed at least two periods before the first period in the sample. The number of periods in each sample is equal to the number of periods in the data compared with the simulations (Argentina 1983–2000). Restricting to samples at least two periods away from the last exclusion period helps avoid extreme observations that may distort the results.[20] The moments reported below correspond to the average across the 400 samples. The behavior of four series is analyzed: the logarithm of income (y), the logarithm of consumption (c), the ratio of the trade balance to output (tb), and the annualized spread (R_s). All series are filtered using the Hodrick-Prescott filter with a smoothing parameter of 1600. Standard deviations are denoted by σ and are reported in percentage terms; correlations are denoted by ρ.

Table 3 reports business cycle moments observed in the data (Argentina 1983–2000) and in Models I and II. With the exception of the debt-to-output

---

[20] In the periods that follow an exclusion period, the government inherits little or no debt from previous periods. The consequence is that in these periods the government borrows a relatively low amount, and thus it pays relatively low spreads compared to what is observed in the remaining observations. In the simulations, these outliers may appear up to two periods after the end of an exclusion period. It is judged that it is more appropriate to present results computed without considering these outliers. Simulation results are contrasted against data from Argentina during a period in which the economy was not excluded from capital markets. Moreover, these outliers can alter the calculations of business cycle statistics (see Hatchondo, Martinez, and Sapriza 2006a).

Table 3 Business Cycle Statistics

	Data	Model I Transitory Shocks	Model II Trend Shocks
$\sigma(y)$	4.08	4.14	4.15
$\sigma(c)$	4.85	4.23	4.38
$\sigma\ (tb)$	1.36	0.20	0.63
$\sigma\ (R_s)$	3.17	0.006	0.013
$\rho\ (c, y)$	0.96	0.99	0.99
$\rho\ (tb, y)$	-0.89	-0.43	-0.29
$\rho\ (R_s, y)$	-0.59	-0.80	-0.06
$\rho\ (R_s, tb)$	0.68	0.85	0.89
Mean Debt Output Ratio (%)	51	6.3	4.8
Rate of Default	75	6.6	24

Notes: The moments correspond to the average across 400 samples. The debt output ratio is measured as the stock of debt divided by the annual output level.

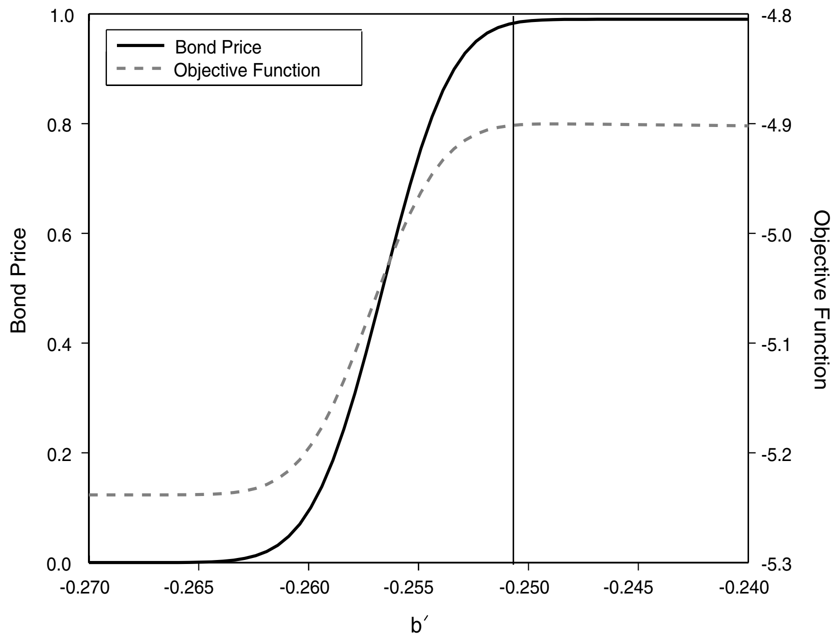
ratio, the business cycle moments for Argentina are taken from Aguiar and Gopinath (2006). The moments are chosen so as to evaluate the ability of the models to replicate the distinctive business cycle properties of emerging economies that were described in the beginning of the article. It must be said that the sample moments for Argentina display the same qualitative features observed in other emerging markets.[21]

The moments in the simulated samples generated by Models I and II are different from the moments reported in Aguiar and Gopinath (2006). The main reason is that they use a different computational method from the one used in this article. While, Aguiar and Gopinath (2006) use a discrete state space method, we use interpolation methods and a nonlinear optimization routine to find the optimal issuance levels. Hatchondo, Martinez, and Sapriza (2006a) demonstrate that the numerical errors incurred by the discrete state space technique may lead to misleading conclusions in some dimensions. The most important one is that the spread volatility becomes negligible once the model is solved using a more accurate method. Other statistics about the behavior of the spread over the business cycle are also susceptible to numerical errors when the model is solved using a discrete state space technique.[22]

[21] As previously mentioned, emerging economies feature interest rates that are high, volatile, and countercyclical; high volatility of consumption relative to income (typically, higher than one); and countercyclical net exports (see, for example, Aguiar and Gopinath 2007; Neumeyer and Perri 2005; and Uribe and Yue 2006).

[22] A second reason behind the discrepancy between the moments in Table 3 and in Aguiar and Gopinath (2006) is that the setups are not exactly the same. In their model, the output loss lasts as long as the exclusion punishment. In the present setup, the output loss lasts for only one period. The Appendix shows that this difference accounts for only a small fraction of the discrepancy in the performance of the two models.

Figure 4 The Effect of the Bond Price Menu on the Objective Function



Notes: Objective function and price function faced by the government when $z = \mu_z$ and b $= -0.25$ (which is within the range observed in the simulations). The vertical line represents the optimal issuance level.

The table shows that both models fail to generate the volatilities of the trade balance and spread observed in the data. In particular, the standard deviations of the spread are two orders of magnitude lower than the value observed in the data. This is an important limitation of both models. Moreover, Models I and II generate 6.6 and 24 defaults in 10,000 periods, respectively. These values are below the ratio of 75 defaults in 10,000 periods computed by Reinhart, Rogoff, and Savastano (2003) using a sample of emerging markets from 1824 to 1999, though it is not clear that this is the frequency that the model should replicate.[23] An alternative procedure is to compare the default rate generated by the model with the default rate implicit in the average spread observed over the sample period (under the assumption of risk-neutral lenders). The value

[23] The model is calibrated to match the macroeconomic behavior of Argentina between 1983 and 2000, while the default rate computed in Reinhart, Rogoff, and Savastano (2003) is based on a different time period and a sample of various countries.

of the latter is 243 defaults for every 10,000 periods, which is even further away from the model predictions.

On the other hand, the models are able to generate a high volatility of consumption relative to income and the sign of the co-movements between the trade balance, spread, and output that are observed in the data.[24]

Discussion of the Results With the Exclusion Punishment

This section describes the tradeoffs the government faces when it decides how much debt to issue. This helps in understanding the logic behind the results presented in Table 3. The section focuses on the model with transitory shocks (z), though the same logic applies to the model with trend shocks.

Given the monotonicity of the default decisions (see Figure 2), the bond price function faced by the government when it decides how much to borrow can be written as

$$q\left(b', z\right) = \frac{1 - F\left(z^*\left(b'\right) \mid z\right)}{1 + r}, \tag{12}$$

where $z^*\left(b'\right)$ denotes the next period endowment shock that makes the government indifferent between defaulting and not defaulting on a debt level b', and F denotes the cumulative distribution function for the next period shock. If the endowment shock in the following period is lower than z^*, the government will default on b'. If it is higher than z^*, the government will pay back b'. In the top panel of Figure 2, $z^*\left(b'\right)$ represents the frontier of the shaded area. Equation (12) shows that the shape of price q mirrors the shape of the probability of observing a default the next period, namely, $F\left(z^*\left(b'\right) \mid z\right)$.

The solid line of Figure 4 displays the bond price menu faced by the government in a period in which the endowment realization is equal to the unconditional mean of the endowment process (μ_z) and the economy is not excluded from capital markets. The sensitivity of the bond price to the issuance volume (b') is given by

$$\frac{q\left(b', z\right)}{\partial b'} = \frac{-f\left(z^*\left(b'\right) \mid z\right)}{1 + r} \frac{\partial z^*\left(b'\right)}{\partial b'}, \tag{13}$$

where f denotes the density function of future shocks. This equation shows that the shape of the bond price depends on two factors: the probability distribution f and the sensitivity of z^* to changes in b' (the shape of the default region). The assumption that future endowment shocks are drawn from a

[24] The table does not report statistics about the current account. Given that both models generate a relatively stable debt level and a low volatility of the interest rate, interest payments also display low volatility. Therefore, the balance of the current account is almost perfectly correlated with the trade balance and inherits the statistical properties of the latter.

Gaussian distribution accounts for the flat portion of the price curve. The thin tails of f explain why the price is almost invariant to $b^{'}$ at issuance volumes such that the threshold z^* takes extreme values.

The bond price plays a central role in understanding the shape of the objective function of the government represented in Figure 4. Formally, the objective function is given by the right-hand side of the Bellman equation:

$$RHS\left(b'\right) = u\left(y\left(1-\lambda\right) + b - q_0\left(b', z\right)b'\right) + \beta \int V(b', z', 0, 0) F_z\left(dz' \mid z\right).$$

For the range of values of b' such that there is no default risk, the present discounted welfare increases with the issuance level, i.e., the burden of starting tomorrow with higher liabilities does not compensate for the extra resources collected for current consumption.[25] As the price per bond starts to fall, there is an extra factor that appears in the tradeoff between current and future consumption: an extra dollar of borrowing implies a lower bond price. In particular, an extra dollar of borrowing implies a decrease in price q received for all the bonds issued in the current period.

If the price function is steep, borrowing an extra dollar is quite costly due to the decrease in bond price received for all bonds issued in the current period. In the stylized model presented in this article, the price function becomes very steep at borrowing levels at which the government pays an interest rate close to the risk-free rate. Consequently, the borrowing levels observed in equilibrium are such that the economy pays low spreads. This explains the low default frequency reported in Table 3.

The top panel of Figure 3 shows that a higher endowment realization enables the government to borrow more without paying a higher spread, but the price function becomes steep at borrowing levels for which the spread is low independently of whether the endowment shock is "low" or "high." This feature contributes to the explanation of why in equilibrium the government chooses to pay low spreads at all endowment realizations, and thus the volatility of the spread is low.

The inability of the model to generate a higher default rate and spread volatility may be a consequence of its simplifying assumptions. Consider, for instance, the assumption that the government can only issue one-period bonds. Recall that as illustrated in Figure 4, the interest rate increases with the borrowing level (the bond price decreases). The assumption of one-period bonds implies that in every period the economy has to roll over its entire stock of debt. Thus, the increase in the interest rate that is due to an extra dollar of borrowing affects the entire stock of bonds and not only the last unit issued. More precisely, consider the decision of whether to borrow $x + 1$ dollars or x

[25] It should be stressed that this result is not general and critically depends on the parameter values chosen, especially the value of the subjective discount factor.

dollars. When the government is renewing its entire stock of debt, the higher interest implied by borrowing $x+1$ instead of x applies to the $x+1$ units issued. This is trivially larger than the cost induced by an increase in the interest rate paid for the last bond issued. This argument illustrates how restricting the government to issue one-period bonds increases the marginal issuance cost, and thus accounts for a fraction of the low spreads generated by the model.

Even though the government is risk averse and lenders are risk neutral, the volatility of consumption is higher than the volatility of output in both models. As discussed in the beginning of Section 4, price q is nondecreasing in the current endowment realization. That is, a higher endowment enables the government to issue more bonds without necessarily paying a higher spread. Given the low value of the discount factor, the economy will seize the opportunity to borrow more whenever it appears, explaining why the economy borrows more in good times. This explains why the model is able to generate a higher volatility of consumption relative to income. Formally, current consumption is determined by current income and net borrowing, namely,

$$c = y - \left(qb' - b\right).$$

Therefore,

$$\sigma^2(c) = \sigma^2(y) + 2\sigma[y - (qb' - b)] + \sigma^2\left(qb' - b\right).$$

The positive covariance between net borrowing $(b - qb')$ and income increases the volatility of consumption relative to income.

Table 3 shows that both models are able to replicate the sign of the comovements between trade balance, spread, and output observed in the data. The fact that the government borrows more in good times leads to a negative correlation between trade balance and output (as observed in the data).[26] The mechanics that determine the sign of the correlation between the spread and output are more complex. On the one hand, if the bond price function faced by the government is kept constant, a higher income realization today reduces the need to borrow, and therefore it reduces the spread that the government is willing to pay for its debt. This generates a negative correlation between income and spread. But the bond price function also changes with the income realization. If the price of the bond becomes less sensitive to the borrowing level at higher income levels, the government would be willing to pay a higher spread at higher income levels. The latter may change the sign of the correlation between the spread and income. For example, the next section shows that

[26] The trade balance is defined as

$$tb = y - c = qb' - b.$$

Thus, the positive correlation between net borrowing and income translates into a negative correlation between trade balance and income.

once the exclusion punishment is eliminated, the spread becomes pro-cyclical in Model II.

5. RESULTS WITHOUT EXCLUSION

This section studies the implications of removing the threat of financial exclusion. Formally, this implies setting the value of ϕ_1 equal to 1. The section describes how the default and saving decisions change when the government cannot be threatened with financial exclusion. This helps to understand how the business cycle statistics change when the exclusion assumption is abandoned, which is discussed later in the section entitled "Business Cycle Properties."

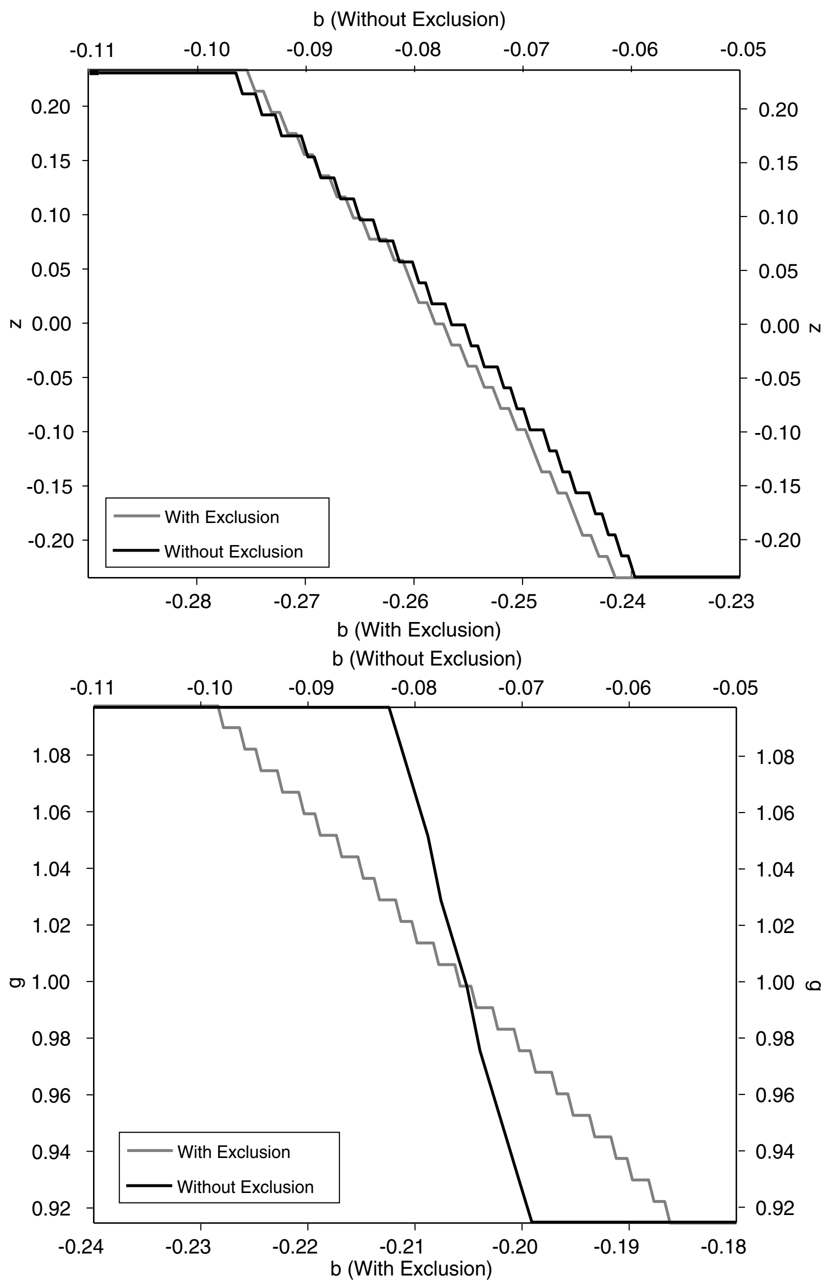
Equilibrium Choices Without the Exclusion Punishment

The common feature across Models I and II is that they are able to sustain a lower debt level when the government does not face the threat of financial exclusion. In other dimensions, the model with transitory shocks without exclusion features higher equilibrium spreads (lower equilibrium issuance prices) and spreads that are more responsive to the endowment shock compared to the model with transitory shocks and exclusion. On the other hand, the model with trend shocks features lower equilibrium spreads and spreads that are less responsive to trend shocks compared to the model with exclusion.

Figure 5 illustrates how the default decisions change when exclusion cannot be used as a punishment in Models I and II. The graphs illustrate that the government defaults at lower debt levels when the threat of financial exclusion is eliminated. The result is not surprising though. The possibility of being excluded from capital markets imposes a cost to every default decision (regardless of the size of the default). Once the threat of financial exclusion is eliminated, it becomes less costly to default, and thus the government finds it optimal to default at lower debt levels.

The top panel of Figure 5 shows that in Model I the default decision becomes relatively more sensitive to the endowment shock (and less sensitive to the debt level) when the threat of exclusion is eliminated. This is due to the fact that the possibility of going into financial autarky is more painful at lower endowment realizations than at higher endowment realizations. When the current shock to the endowment level is higher, the need to smooth out consumption by borrowing is weaker, and therefore the value assigned to retaining access to capital markets is lower. In other words, the government would suffer less from being excluded from capital markets if its endowment level is higher. When the threat of financial exclusion is eliminated, the overall cost of defaulting decreases more at lower endowment realizations than at

Figure 5 Default Regions With and Without Exclusion



Notes: Endowment shocks at which the government is indifferent between defaulting and not defaulting in the model with *transitory* shocks (top panel) and in the model with *trend* shocks (bottom panel). The scale on the bottom corresponds to the models *with exclusion*. The scale on the top corresponds to the models *without exclusion*.

higher endowment realizations. This accounts for the flatter default region in the top panel of Figure 5.

The picture looks different in the model with trend shocks. The bottom panel of Figure 5 shows that the default region becomes steeper without exclusion. In this case, the mechanism described in the previous paragraph is also present, but there is an additional effect. A higher growth rate in the current period not only means that there are more resources available for current consumption but also that future growth rates are likely to be high—recall that there is persistence in growth rates. Consequently, unlike a higher transitory shock, a higher growth shock introduces an incentive to borrow more on account of the future increases in the endowment. This means that the value assigned to retaining access to capital markets is larger when the current growth rate is higher. When the threat of financial exclusion is eliminated, the overall cost of defaulting decreases more at higher endowment realizations than at lower endowment realizations. This explains the change in the slope of the default regions in the model with trend shocks.

The change in the shape of the default region plays an important role in understanding the change in the shape of the price function. The formal link between the two is described in equation 13. For example, the steeper the default region (the higher the expression $\left|\frac{\partial z^*(b')}{\partial b'}\right|$), the steeper the price function.

Figure 6 shows the price functions faced by the economy in Model I with and without exclusion.[27] The charts in Figure 6 show that when the threat of exclusion is eliminated, the bond price starts to decrease at a lower issuance level. This mirrors the shift of the default region due to a lower cost of default. The graphs show that the moderate change in the slope of the default regions observed in the top panel of Figure 5 translate into a moderate change in the slope of the price functions.

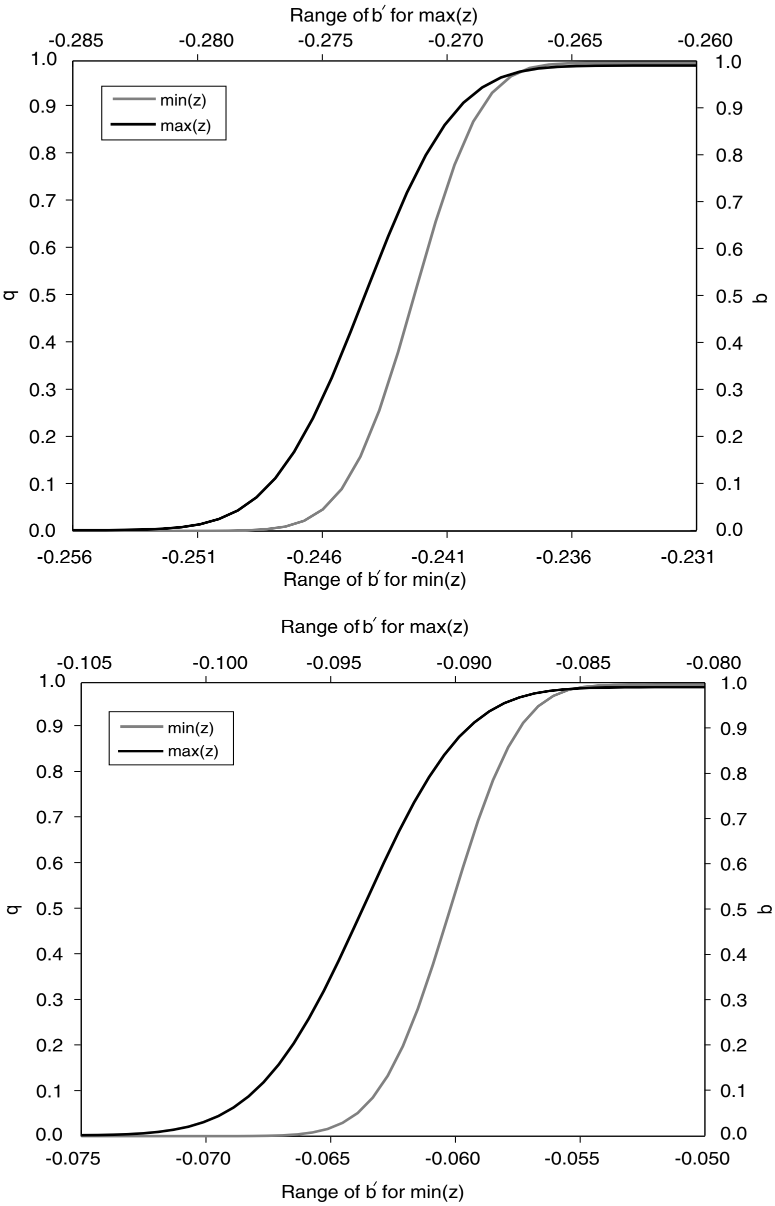
Figure 7 shows the price function faced by the economy in Model II with and without exclusion.[28] The charts show that the steeper default regions that are observed when the threat of financial exclusion is eliminated translate into steeper price functions.

---

[27] Notice that the top panel of Figure 6 represents the same functions as the top panel of Figure 3 but with different scales. The graph is reproduced again in order to facilitate the comparison of the shape of the price functions in the models with and without exclusion.

[28] The top panel of Figure 6 represents the same functions as the bottom panel of Figure 3 but with different scales.

Figure 6 Bond Price Menus in Model I With and Without Exclusion



Notes: Bond price as a function of the issuance level in the model with *transitory* shocks and *exclusion* (top panel) and in the model with *transitory* shocks and *without exclusion* (bottom panel). The high (low) shock is three standard deviations higher (lower) than the unconditional mean of z. The scale of bond issuances when the endowment shock is low is described by the bottom horizontal axes. Likewise, the scale of bond issuances when the endowment shock is high is described by the top horizontal axes. The different scales used in the horizontal axes facilitate the comparison of the shape of the price functions.

Figure 8 displays the equilibrium issuance price in Model I as a function of the endowment shock realization in the specifications with and without exclusion. The graphs show that the price at which the government issues debt is lower and more sensitive to the endowment shock in the setup without exclusion.

Figure 9 shows the bond prices paid in equilibrium in the model with shocks to the trend with and without exclusion. The graphs show that the prices at which the government issues debt are higher in the setup without exclusion. The correlation with the endowment shock also changes. When the government can be threatened with financial exclusion the spread decreases with respect to the shock to the trend. But when the government cannot be threatened with financial exclusion, the spread increases with respect to the shock to the trend.[29]

The Mechanics of the Equilibrium Behavior of the Spread

This section discusses the differential behavior of the spread in Models I and II once the exclusion assumption is abandoned. Consider first the Euler equation that determines the optimal borrowing level

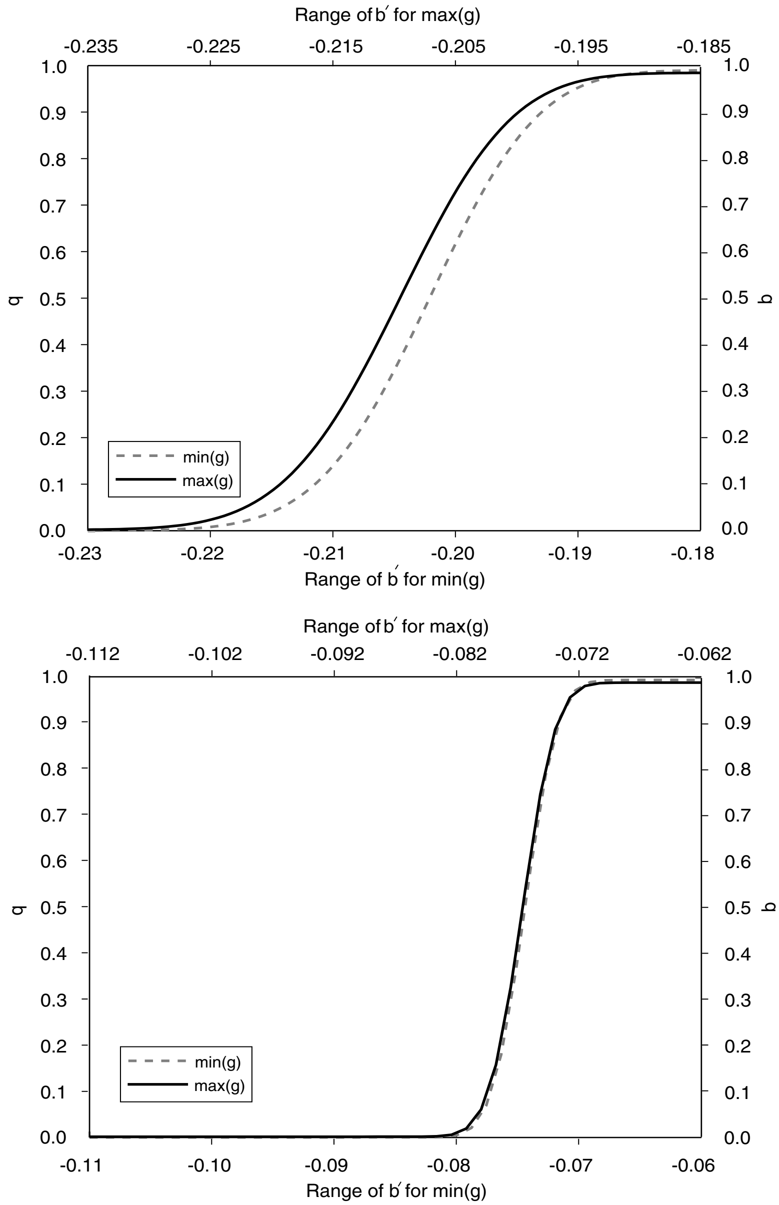
$$u'(c)\, q_0(b', z, \Gamma, g) = \left\{ \begin{array}{c} \beta \int \int \frac{\partial V(b',z',\Gamma',g',h)}{\partial b'} F(dz' \mid z) F(dg' \mid g) - \\ u'(c) b' \frac{\partial q_0(b',z,\Gamma,g)}{\partial b'}. \end{array} \right\} \quad (14)$$

The left-hand side of the equation captures the marginal benefit of issuing one more bond today, i.e., the increase in current consumption. The right-hand side captures the marginal costs. The first term represents the "future marginal cost." Issuing one more unit of debt today makes the economy poorer in the future—the government will either have to pay back its debt or face the cost of defaulting. The second term on the right-hand side represents the "present marginal cost." This is the cost derived from decreasing the price of all the bonds issued today. The role of the latter was discussed more extensively in the end of Section 4.

In both models the optimal issuance volume is lower when the government does not face the threat of default. An immediate consequence is that it depresses the present marginal cost (the second term in the right-hand side of equation (14) is the product of the borrowing level times the sensitivity of the price to the borrowing level). The decrease in the marginal cost induced by a lower borrowing level may be compensated, in part, by accepting a lower price for each bond issued, which reduces the marginal benefit of borrowing. This can explain the lower equilibrium bond prices (higher spread) that are observed in Model I when the threat of exclusion is eliminated (see Figure 8).

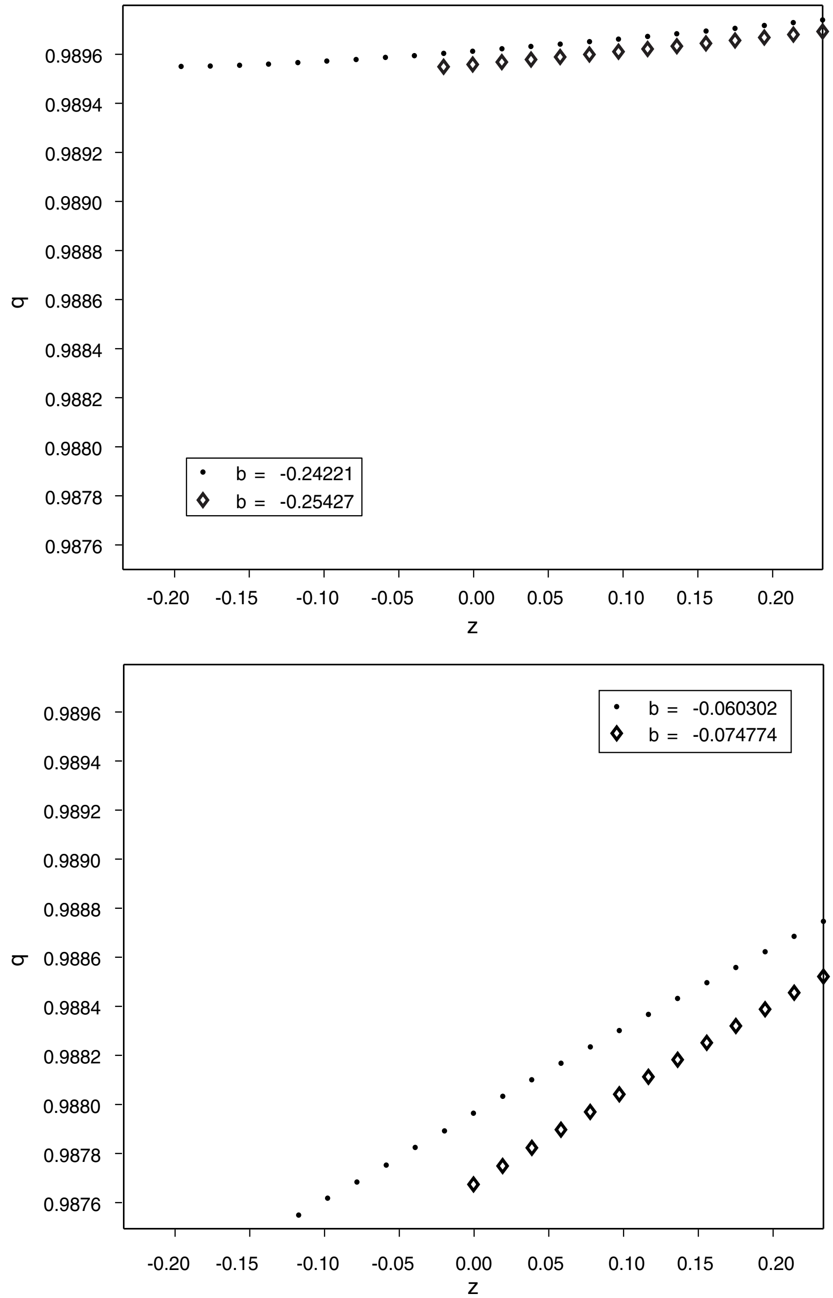
[29] The beginning of Section 5 provides some intuition for the differential behavior of the spread displayed by Models I and II once the exclusion assumption is abandoned.

Figure 7 Bond Price Menus in Model II With and Without Exclusion



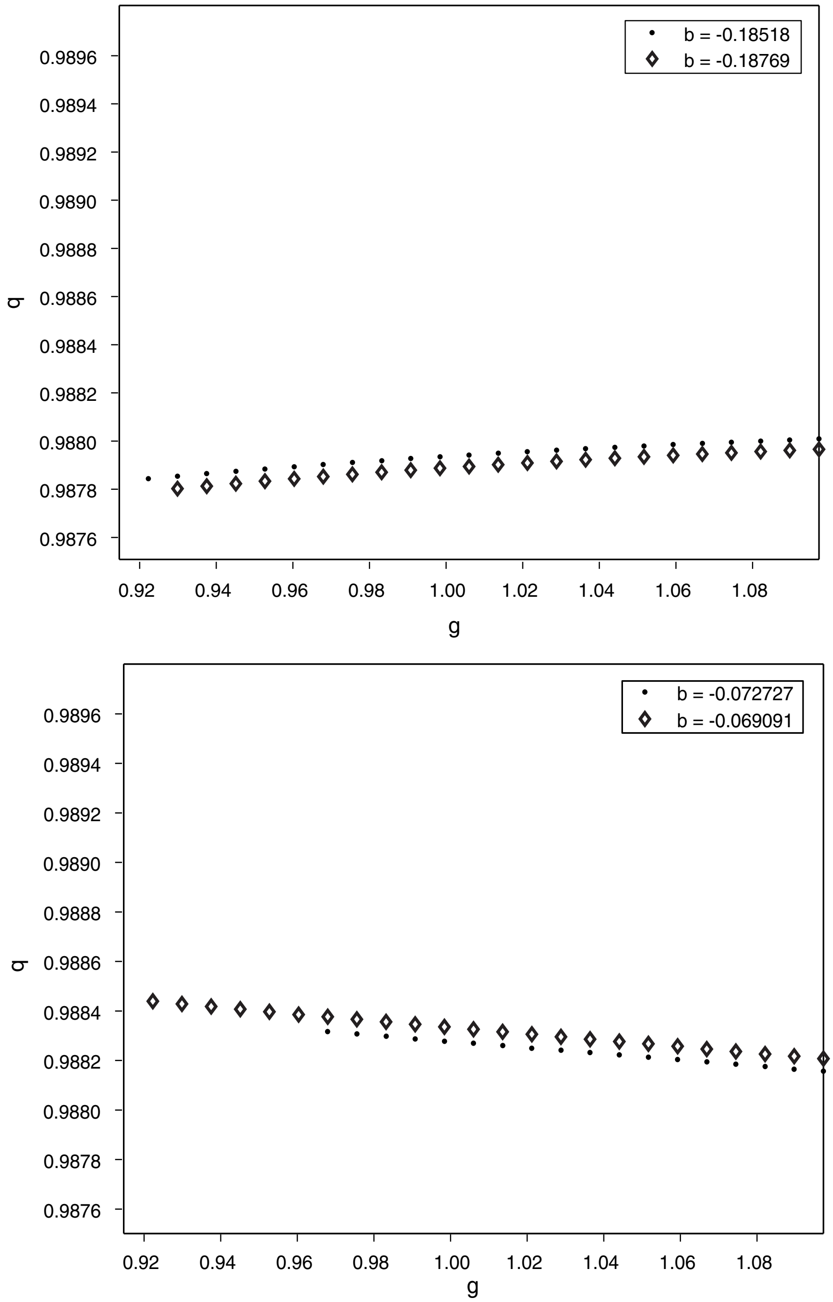
Notes: Bond price as a function of the issuance level in the model with *trend* shocks and *exclusion* (top panel) and in the model with *trend* shocks and *without exclusion* (bottom panel). The high (low) shock is three standard deviations higher (lower) than the unconditional mean of g. The scale of bond issuances when the trend shock is low is described by the bottom horizontal axes. Likewise, the scale of bond issuances when the trend shock is high is described by the top horizontal axes. The different scales used in the horizontal axes facilitate the comparison of the shape of the price functions.

Figure 8 Bond Prices Observed in Equilibrium in Model I



Notes: Bond price observed in equilibrium in the model with *transitory* shocks and *exclusion* (top panel) and in the model with *transitory* shocks and *without exclusion* (bottom panel).

Figure 9 Bond Prices Observed in Equilibrium in Model II



Notes: Bond price observed in equilibrium in the model with *trend* shocks and *exclusion* (top panel) and in the model with *trend* shocks and *without exclusion* (bottom panel).

Table 4 Business Cycle Statistics Computed With and Without Exclusion Punishment

		Transitory Shocks		Trend Shocks	
	Data	**With Exclusion**	**Without Exclusion**	**With Exclusion**	**Without Exclusion**
$\sigma(y)$	4.08	4.14	4.10	4.15	4.16
$\sigma(c)$	4.85	4.23	4.19	4.38	4.23
$\sigma\ (TB/Y)$	1.36	0.20	0.21	0.63	0.23
$\sigma\ (R_s)$	3.17	0.006	0.05	0.013	0.015
$\rho\ (c, y)$	0.96	0.99	0.99	0.99	0.99
$\rho\ (TB/Y, y)$	-0.89	-0.43	-0.43	-0.29	-0.29
$\rho\ (R_s, y)$	-0.59	-0.81	-0.95	-0.06	0.40
$\rho\ (R_s, TB/Y)$	0.68	0.85	0.69	0.89	-0.96
Mean Debt Output Ratio (%)	51	6.3	1.7	4.8	1.8
Rate of Default	75	6.6	25	24	20

Notes: The debt output ratio is measured as the stock of debt divided by the annual output level.

A similar effect is present in Model II. But in this case the bond price becomes steeper in the setup without exclusion. This effect alone tends to increase the present marginal cost of borrowing, and therefore unlike Model I, a lower bond price is not necessary to satisfy equation (14). This can explain why the levels of the equilibrium bond prices in Model II do not change significantly when the threat of exclusion is eliminated (see Figure 9).

The forces behind the changes in the slope of the spread with respect to the endowment shock are more difficult to tease out. As explained in the end of Section 4, the equilibrium relationship between the spread and the endowment shock depends on various effects, and the sign of the relationship does not necessarily need to be negative. In fact, Figure 9 shows that in the model with trend shocks and no exclusion, the spread increases with the growth shock.

Business Cycle Properties Under No Exclusion

The business cycle statistics reported in Table 3 are recalculated for an economy without the exclusion punishment and presented in Table 4 (the statistics in Table 3 are reproduced in order to facilitate comparison).

The implications of removing the exclusion punishment reported in Table 4 are consistent with the discussion of equilibrium choices in the beginning of Section 5. Table 4 shows that the assumption of exogenous exclusion is responsible for a high fraction of the debt level supported in the model with exclusion. Recall that in this model the government chooses borrowing levels that allow the government to pay very low spreads. These levels are lower when the default cost raised by the threat of financial exclusion is removed.

Table 4 shows that both models quantitatively fail along important dimensions with or without the assumption of financial exclusion: the default rate, the volatility of the spread, and the volatility of the trade balance are too low compared with the data. Even though the overall performance is poor, the behavior of the model with transitory shocks shows a moderate improvement, while the behavior of the model with trend shocks deteriorates when the exclusion assumption is eliminated. The model with endowment shocks and no exclusion displays a higher default rate and spread volatility compared to the model with exclusion (but still far below the data), while the remaining business cycle statistics are not substantially different. On the other hand, the sign of the correlation between output and the spread, and between the spread and the trade balance are reversed and become counterfactual in the model with trend shocks and no exclusion.

The higher default rate generated by Model I when the threat of financial exclusion is eliminated is consistent with the higher equilibrium spread described in Figure 8. The higher spread volatility observed in the setup without exclusion is also consistent with Figure 8, which shows a higher sensitivity of the spread with respect to output in the model without exclusion. The lower default rate and similar spread volatility generated by Model II when the threat of financial exclusion is eliminated are consistent with the adjustments illustrated in Figure 9.

6. CONCLUDING REMARKS

This article discusses the quantitative performance of sovereign default models and explains how the performance is affected by the assumption that countries can be exogenously excluded from capital markets after a default. The article compares the performance of a stylized model with and without the threat of exclusion. It is shown that the exclusion assumption explains a high fraction of the sovereign debt that can be sustained in equilibrium but does not significantly alter the remaining business cycle statistics of the model. In effect, the model without exclusion generates annual debt-output ratios of less than 2 percent. The model with exclusion generates annual debt-output ratios of 4.8 percent when the shocks hit the growth rate and of 6.3 percent when the shocks hit the endowment level. The article shows that in the model with shocks to the endowment level, the default decision becomes slightly more sensitive to the endowment shock and, therefore, less sensitive to the debt level. This helps reduce the sensitivity of the bond price to the borrowing level. On the other hand, in the model with shocks to the trend the default decision becomes relatively less sensitive to the endowment shock, which increases the sensitivity of the bond price to the borrowing level. Given that in this class of models the excessive sensitivity of the default probability to the borrowing level limits the models' quantitative performance, the previous effects may help explain why the performance of the model with shocks to the endowment level shows moderate improvement and why the performance of the model with shocks to the trend deteriorates. In spite of this, both models still fail along important dimensions. The default rate, the volatilities of the trade balance and of the spread, and the debt levels are too low compared to the data. These shortcomings suggest that the exclusion assumption does not play an important role, and therefore future studies that do not rely on the threat of financial exclusion will not necessarily be handicapped in explaining the business cycle in emerging economies. These shortcomings also suggest that other assumptions of the model must be modified in order to bring the model closer to the data.

APPENDIX: THE MODEL WITH STOCHASTIC DURATION OF OUTPUT LOSS

As mentioned before, the model introduced in Section 2 does not exactly coincide with the model presented in Aguiar and Gopinath (2006). They assume that following a default episode, the duration of lower output lasts as long as the time of exclusion. Table 5 shows that the business cycle statistics presented in Table 3 are not greatly affected by the choice of the process of output loss—statistics computed with a stochastic duration of output loss are taken from Hatchondo, Martinez, and Sapriza (2006a) who solve the model in Aguiar and Gopinath (2006) with the computational method used in this article.

Table 5 Business Cycles Under Different Specifications of the Output Loss

	Transitory Shocks		**Trend Shocks**	
	One-Period Output Loss	**Stochastic Duration of Output Loss**	**One-Period Output Loss**	**Stochastic Duration of Output Loss**
$\sigma(y)$	4.14	4.13	4.16	4.16
$\sigma(c)$	4.23	4.25	4.38	4.39
$\sigma\ (TB/Y)$	0.20	0.24	0.63	0.65
$\sigma\ (R_s)$	0.006	0.007	0.013	0.015
$\rho\ (c, y)$	0.99	0.99	0.99	0.99
$\rho\ (TB/Y, y)$	-0.43	-0.43	-0.29	-0.29
$\rho\ (R_s, y)$	-0.81	-0.74	-0.06	-0.09
$\rho\ (R_s, TB/Y)$	0.85	0.89	0.89	0.91
Mean Debt Output Ratio (%)	6.8	6.8	4.7	4.8
Rate of Default	6.6	7.8	24	22

Notes: Business cycle statistics from models with output loss in one period and models with a stochastic duration of the output loss. The debt output ratio is measured as the stock of debt divided by the annual output level.

REFERENCES

Aguiar, Mark, and Gita Gopinath. 2006. “Defaultable Debt, Interest Rates and the Current Account.” *Journal of International Economics* 69 (1): 64–83.

Aguiar, Mark, and Gita Gopinath. 2007. “Emerging Markets Business Cycles: The Cycle is the Trend.” *Journal of Political Economy* 115 (1): 69–102.

Arellano, Cristina. Forthcoming. “Default Risk and Income Fluctuations in Emerging Economies.” *American Economic Review.*

Arellano, Cristina, and Ananth Ramanarayanan. 2006. “Default and the Term Structure in Sovereign Bonds.” Working Paper, University of Minnesota.

Athreya, Kartik B., and Hubert P. Janicki. 2006. “Credit Exclusion in Quantitative Models of Bankruptcy: Does It Matter?” Federal Reserve Bank of Richmond *Economic Quarterly* 92 (1): 17–50.

Bai, Yan, and Jing Zhang. 2006. “Financial Integration and International Risk Sharing.” Working Paper, University of Michigan.

Borensztein, Eduardo, and Paolo Mauro. 2004. “The Case for GDP-Indexed Bonds.” *Economic Policy* 19 (38): 165–216.

Chuhan, Punam, and Federico Sturzenegger. 2005. “Defaults Episodes in the 1980s and 1990s: What Have We Learned?” In *Managing Economic Volatility and Crises*, eds. Joshua Aizenman and Brian Pinto, New York, NY: Cambridge University Press: 471–519.

Cole, Harold L., James Dow, and William B. English. 1995. “Default, Settlement, and Signalling: Lending Resumption in a Reputational Model of Sovereign Debt.” *International Economic Review* 36 (2): 365–85.

Cuadra, Gabriel, and Horacio Sapriza. 2006a. “Sovereign Default, Interest Rates and Political Uncertainty in Emerging Markets.” Banco de México, Working Paper 2006-02.

Cuadra, Gabriel, and Horacio Sapriza. 2006b. “Sovereign Default, Terms of Trade, and Interest Rates in Emerging Markets.” Banco de México,Working Paper 2006-01.

Eaton, Jonathan, and Mark Gersovitz. 1981. “Debt With Potential Repudiation: Theoretical and Empirical Analysis.” *Review of Economic Studies* 48 (2): 289–309.

Eichengreen, Barry, and Richard Portes. 2000. “Debt Restructuring With and Without the IMF.” Manuscript, University of California, Berkeley.

Gelos, Gaston, Ratna Sahay, and Guido Sandleris. 2004. "Sovereign Borrowing by Developing Countries: What Determines Market Access?" IMF Working Paper 04/221.

Hatchondo, Juan Carlos, Leonardo Martinez, and Horacio Sapriza. 2006a. "Computing Business Cycles in Emerging Economy Models." Federal Reserve Bank of Richmond, Working Paper No. 06-11.

Hatchondo, Juan Carlos, Leonardo Martinez, and Horacio Sapriza. 2006b. "Heterogeneous Borrowers in Quantitative Models of Sovereign Default." Federal Reserve Bank of Richmond, Working Paper No. 07-1.

IMF. 2002. "Sovereign Debt Restructurings and the Domestic Economy Experience in Four Recent Cases." Policy Development and Review Department, February 21.

Kumhof, Michael. 2004. "Fiscal Crisis Resolution: Taxation Versus Inflation." Manuscript, Stanford University.

Kumhof, Michael, and Evan Tanner. 2005. "Government Debt: A Key Role in Financial Intermediation." IMF Working Paper 05/57.

Lizarazo, Sandra. 2005a. "Sovereign Risk and Risk Averse International Investors." ITAM, Working Paper.

Lizarazo, Sandra. 2005b. "Contagion of Financial Crises in Sovereign Debt Markets." ITAM, Working Paper.

Meyersson, Erik. 2006. "Debt Intolerance and Institutions, an Empirical Investigation into the Institutional Effect of Sovereign Debt." Mimeo, Institute for International Economic Studies, Stockholm University.

Neumeyer, Pablo A., and Fabrizio Perri. 2005. "Business Cycles in Emerging Economies: The Role of Interest Rates." *Journal of Monetary Economics* 52 (2): 345–80.

Reinhart, Carmen M., Kenneth S. Rogoff, and Miguel A. Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity* 1: 1–62.

Schmitt-Grohé, Stephanie, and Martin Uribe. 2003. "Closing Small Open Economy Models." *Journal of International Economics* 61 (1): 163–85.

Sturzenegger, Federico, and Jeromin Zettelmeyer. 2005. "Has the Legal Threat to Sovereign Debt Restructuring Become Real?" Working Paper, Harvard University.

Uribe, Martin, and Vivian Yue. 2006. "Country Spreads and Emerging Countries: Who Drives Whom?" *Journal of International Economics* 69 (1): 6–36.

Yue, Vivian. 2005. "Sovereign Default and Debt Renegotiation." Working Paper, New York University.

Economic Quarterly—Volume 93, Number 3—Summer 2007—Pages 287–315

Barriers to Foreign Direct Investment Under Political Instability

Marina Azzimonti and Pierre-Daniel G. Sarte

Foreign direct investment (FDI), as pointed out by Kindleberger (1969), arises when the host country has an investment opportunity that it cannot exploit by itself because it lacks the means or technical know-how, or because of market incompleteness (that is, access to capital markets is restricted). A multinational corporation (MNC) may be able to exploit such an opportunity because it has the necessary capital, technology, and managerial skills to do so. Even though the return to foreign direct investment is potentially large in many developing countries (for example, the opening up of Eastern Europe provided advantages to multinational firms because of the low cost of labor, low levels of capital in place, and the proximity to major markets), the flow of direct investment is concentrated in just a few countries.[1] Lucas (1990) attributes this lack of FDI in countries with potentially large marginal returns to capital to the fact that many developing countries face higher political risk than industrialized ones.

A distinctive characteristic of FDI is that once an investment has been made, a foreign investor cannot prevent the government in the host country from changing the environment in which the investment decision was made. Despite attempts to establish international tribunals, contracts between multinational corporations (MNCs) and sovereign countries are almost impossible to enforce. The quality of institutions, and in particular, the degree of protection of property rights, are key in determining the expected return to foreign

We would like to thank Andreas Hornstein, Leonardo Martinez, Borys Grochulski, and Nashat Moin for their comments. Also, we would like to thank Kevin Bryan for excellent research assistance. The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.
E-mails: marina-azzimonti@uiowa.edu and pierre.sarte@rich.frb.org.

[1] The United Nations (1996) reports that 80 percent of the total investment flowing to developing countries in 1995 was received by only ten countries.

investors. Countries with relatively poor legal protection of assets, and a high degree of political instability, generally exhibit high rates of expropriation and this makes investment less attractive. In practice, expropriation can take different forms. A *direct* act of expropriation involves nationalization of foreign-owned corporations, in which the government simply takes control of the capital stock (Kobrin 1980, 1984). There are also *indirect* forms of expropriation that multinational corporations face. Examples include excessive taxation, capital controls, manipulation of exchange rates, and bribes and permits demanded by government officials.

In this article, we describe some stylized facts about expropriation episodes and other lessons learned from the empirical literature on FDI. We then summarize some of the main theories attempting to explain the effects of expropriation on investment and growth. Finally, we develop a theory that relates each type of expropriation to political instability and concentration of power.

A simple two-period political economy model is presented in which groups with access to an expropriation technology alternate in power according to an exogenous probability. The group that controls the government in the first period has the ability to obtain bribes from foreign investors who are attempting to gain access to production in the host country. This form of indirect expropriation is analogous to an investment tax, in the sense that it distorts the optimal allocation of international capital by imposing additional costs to potential investors. After investment decisions have been made (in the second period), the group in office decides how much capital should be seized or nationalized, a direct form of expropriation.

Following the literature on FDI, we will assume that any capital expropriated by the host country becomes unproductive. This stylized representation tries to characterize the empirical observation that MNCs are usually more efficient in running production than the host country. For example, Minor (1994) documents that about 35 percent of all enterprises that were expropriated between 1960 and 1979 have been privatized between 1980 and 1992, indicating public "disillusion with the typical result of expropriation, the state-owned enterprise" (see Biais and Perotti 2002 for more on recent trends on privatizations). Theoretically, the costs associated with expropriation arise mainly because of two reasons. First, there is usually a reduction in the technological spillovers embodied in foreign capital. Second, because the capital installed by foreign investors may be specific to the manager's skills, it may take time for domestic workers to acquire the know-how needed to operate the foreign technology. As a consequence, reductions in the capital stock installed by MNCs imply productivity losses and depressed domestic wages.

At any point in time, the benefit associated with expropriation is given by the amount of goods that can be transferred from MNCs to domestic agents. The tradeoff faced by policymakers is, therefore, given by the redistributive gains of expropriation versus the income loss suffered by local workers.

A key assumption is that there are no institutional barriers to discretionary redistribution, so any group can appropriate all expropriated resources. Because the group in power is not forced to transfer resources to other groups, a "tragedy of the commons" arises: there is too much expropriation in equilibrium. A tragedy of commons occurs when property rights of an asset cannot be enforced; a typical example is fishing on a lake. Typically this gives rise to over-consumption or under-investment (see Gordon 1954 or Lancaster 1973). In our model, it is precisely the fact that groups cannot ensure ex ante that they will receive the benefits of expropriation in the future that cause over-expropriation in the first period, making the level of bribes inefficiently large. The degree of such inefficiency is related to how likely it is that the current group in government retains power in the second period. That is, the degree of such inefficiency is related to the political instability.

While countries that have higher political instability are predicted to exhibit higher levels of indirect expropriation, direct expropriation levels are lower. The intuition is as follows: because each group finds its chances of being in power in the second period very unlikely, it becomes shortsighted and demands a large quantity of bribes when in power (i.e., in the first period). This discourages investment, and the reduction of capital decreases the marginal cost of direct expropriation, encouraging more expropriation in the second period. The marginal benefit is reduced as well because the tax base shrinks, which reduces the incentives to expropriate. Under a Cobb-Douglas technology assumption, the latter effect dominates and direct expropriation goes down.

A second interesting result derived in this article is related to the concentration of power. Following Tornell and Lane (1999), power is concentrated when there are few groups competing for government. They find that the relation between indirect expropriation and the number of groups in power is non-monotonic. When there is high concentration of power initially, a dilution of concentration results in more indirect expropriation, but this relationship reverts when concentration is small (i.e., there is a large number of groups to begin with). Direct expropriation, on the contrary, always increases with the number of groups. We provide some details on the intuition behind this result at the end of Section 4.

The organization of the paper is as follows. We define the different types of expropriation in Section 1 and summarize the main empirical findings in the literature in Section 2. We then proceed to describe some of the most influential theoretical articles on expropriation in Section 3. In Section 4, our model is described and the main results are derived. Section 5 concludes.

1. DEFINING EXPROPRIATION

Expropriation refers, in general, to policies that adversely affect the private value of the stock and/or returns of foreign investment. As mentioned in the previous section, we can distinguish between "direct" and "indirect" expropriation.

OECD (2004) provides an extensive analysis of the concept of expropriation, where jurisprudence, state practice, and literature on international investment law are considered. According to the survey, *direct expropriation* is "... an act where there is a compulsory transfer of property rights by the host state.... An investment is nationalised or otherwise directly expropriated through formal transfer of title or outright physical seizure. In addition to the term expropriation, terms such as 'dispossession,' 'taking,' 'deprivation,'or privation' are also used." Kobrin (1980, 1984) and Minor (1994) define direct expropriation as the "forced divestment of equity ownership of a foreign direct investor." The principal characteristic is that such divestment is involuntary, against the will of the owners and/or managers of the enterprise, and must entail managerial control through equity ownership across national borders.

Indirect expropriation stands for other forms of change in the institutional environment that reduce the value of an investment, but in which property is not necessarily seized. Schlemmer-Schulte (1999) characterizes indirect expropriation "... as excessive and repetitive tax or regulatory measures that have a *de facto* confiscatory effect in that their combined results deprive the investor in fact of his ownership, control or interests in the investment ..." This may be accomplished, in addition to the raising of taxes, through manipulation of exchange rates (i.e., devaluations), fees or bribes charged to the enterprise, the return of the firm to public ownership at unfair terms, the stiffening of regulation, or the institution of non-tariff barriers, such as restrictions in the repatriation of profits or other capital transactions (referred to as "transfer risk" by insurance companies). This form of indirect expropriation is also referred to as "disguised" or "creeping expropriation." In contrast to the case of direct expropriation, there is no generally accepted definition of indirect expropriation in international law. Moreover, the distinction between this form of expropriation and non-compensable regulation (i.e., antitrust laws, environmental protection, etc.) is not clear.[2]

2. LESSONS FROM THE EMPIRICAL LITERATURE

In this section, we will analyze alternative forms of expropriation and describe their changes over the past 30 years. Afterward, we summarize some of the

[2] See Organization for Economic Cooperation and Development (2004) for an extensive discussion on the issue.

empirical articles documenting the relationship between expropriation (and other measures of the quality of institutions) and FDI.

Direct Expropriation

According to Minor (1994), there were 575 expropriation acts between 1960 and 1992, committed by 79 developing host countries against foreign multinationals. Africa was the region with the highest concentration of expropriation events in the 1960s and 1970s, but Latin America and Asia became more active during the 1980s. The manufacturing and petroleum sectors were the most affected by direct expropriation: they account for about 40 percent of all expropriation events between 1960 and 1964, and this percentage rises to almost 50 percent in the period 1976–1979. Jensen (2005) points out that another industry recently affected by major political events was privately financed infrastructure, in which some projects have been directly expropriated (for example, the government of Thailand's seizure of a private expressway in 1993). Li (2004) documents that out of 520 expropriation acts committed between 1960 and 1990, autocratic governments committed 423 acts while democratic governments committed only 97 acts. This finding relates to the fact that democratic governments have stronger institutions protecting property rights.

Minor shows a decline in the number of expropriation events after the 1970s. This is explained by the fact that international conditions in the late 1970s increased the benefits of FDI inflows and the freedom of action over some multinational corporations was limited. For example, in 1990 a paragraph in the Chinese-Foreign Joint Venture Law added a "no nationalization" clause (Robertson and Chen 1990). Tanzania adopted the National Investment Protection Policy Act that offers legal protection against nationalization (Corkran 1991). Of course, whether host countries respect such agreements ex post is not obvious. A more important factor that reduced the incentives to nationalize multinational corporations was the failure of state-owned enterprises. As mentioned earlier, more than 35 percent of the enterprises that were expropriated prior to 1980 were subsequently privatized. This indicates that multinational corporations have an advantage over domestic governments in running production (because investments are specific to the skills of their foreign managers, for example), an assumption that will be made in the theoretical section of the article.

According to the Organization for Economic Cooperation and Development (OECD), "Disputes on direct expropriation—mainly related to nationalization that marked the 1970s and 1980s—have been replaced by disputes related to foreign investment regulation and 'indirect expropriation' " (OECD 2004, 2). The following section describes the particular form that this type of expropriation has taken in recent years.

Indirect Expropriation

Indirect expropriation acts are more difficult to document in a consistent manner because of the lack of a formal or legal definition. In this subsection, we will restrict attention to a set of examples to highlight the nature of these expropriation acts.

Argentina's financial crisis of 2001–2002, when the "corralito" was imposed, provides a good example of indirect expropriation: the government restricted capital transactions and "pesified" contracts and financial assets. Foreign firms' funds were converted into pesos, and many contracts, especially in infrastructure, were rewritten or canceled. At the same time, capital was not allowed to leave the country (hence the name, "corralito," which means "little fence").

Janeba (2002) provides some other examples of indirect expropriation. In 1995, China announced the dissolution of various benefits that foreign firms received in the form of exemptions from custom duties or tax rebates when using local materials. Russia frequently considered introducing a "super profits tax" for foreign oil companies investing in Russia. Government renegotiation of power, electricity, and water contracts after financial crises in Argentina, Indonesia, Pakistan, and the Philippines constitute further examples (see Moran 2003). More recent examples include foreign oil companies being forced out of their joint venture contracts, for example, such as the company, TNK-BP in Russia.

Shleifer and Vishny (1993) argue that indirect expropriation is particularly distortive for countries with unstable governments in which an entrepreneur may have to bribe several public officials and still face the possibility that none of them really has the power to allow the project to proceed.

Stylized Facts

Trends

Researchers at the World Bank's Multilateral Investment Guarantee Agency (MIGA) found that U.S. investors in emerging markets were subject to both direct and indirect acts of expropriations between 1970 and 2001. The researchers note that between 1971 and 1980, U.S. investors were exposed to restrictions on transferring and repatriating funds (transfer risk) and also subject to a number of direct expropriations. During the period of 1981–1990, an even greater increase in the number of transfer risks claims as well as major reductions in the number of expropriations occurred. Chifor (2002) notes, "In the past two decades, indirect expropriation has supplanted direct takings as the dominant form of state interference with foreign investment, as host countries have learned that more value can be extracted from foreign enterprises through the more subtle instrument of regulatory control rather than outright seizures." The period of 1996–2000 was risky for multinational corporations, mainly because political violence and civil war claims increased dramatically. For most firms, however, direct expropriation was the most damaging. As Jensen (2005) notes, "Of all the dollars paid out by OPIC from 1970-1978, 96% of these claims were for expropriation. From 1991-2004, even after the major financial crises that triggered a number of transfer claims, 84% of the settlement amounts of OPIC claims were for expropriation."[3]

FDI, Expropriation, and Institutions

There are a large number of empirical articles that attempt to assess the quantitative importance of expropriations and the quality of institutions on FDI inflows. Most studies make no distinction between the effects of direct and indirect forms of expropriation. An exception to this are articles focusing on corruption, a form of creeping expropriation. Mauro (1995) finds that corruption has a negative effect on total and private investment, thus hindering growth. Wei (2000), using data on OECD countries, shows that corruption indices are strongly and negatively correlated with FDI inflows. For example, he estimates that an increase in Singapore's level of corruption to that of Mexico's would have the same negative effect on inward FDI as raising the tax rate on multinational corporations by 50 percentage points. Hines (1995) documents a reduction in U.S. FDI in the period following the 1977 U.S. Foreign Corrupt Practices Act, which stipulated penalties for U.S. multinational firms found to be bribing foreign officials. Asiedu (2006), using a panel data for 22 countries over the period 1984–2000, shows that a decline in Nigeria's level of corruption to that of South Africa's has the same positive effect on FDI as increasing the share of fuels and minerals in total exports by about 35 percent. He concludes that countries that are small or lack natural resources can attract FDI by improving their institutions and policy environment.

Variables contained in the Political Risk Services/International Country Risk Guide (PRS/ICRG) political risk dataset, such as corruption in government, expropriation risk, bureaucratic quality, risk of repudiation of contracts by the government, and law and order, are used in other studies to explain differences in FDI inflows across countries. These variables are collected in order to provide a comparable measure across countries of how expected returns to capital investment are reduced by direct and indirect forms of expropriation. While some components such as expropriation risk, for example, only incorporate the probability that capital is expropriated after investment, others such as corruption in government, for example, refer to reductions in profitability that will occur almost with certainty (i.e., bribes).

[3] The United States Overseas Private Investment Corporation (OPIC) provides investment insurance for U.S. firms.

Daude and Stein (2001), using a simple average of the variables in the PRS/ICRG dataset mentioned previously (for the year 1995), find that a one standard deviation improvement in the quality of institutions increases FDI by a factor of 2.2. When focusing on risk of repudiation of contracts by the government, an improvement of one standard deviation—for example, from the level of Egypt to that of Finland—increases FDI by a factor of 1.4. They also find that variables measuring economic policy predictability are positively correlated with FDI inflows. Busse and Hefeker (forthcoming), using the same dataset for the period 1984–2005, find that the quality of institutions is a relevant factor for determining FDI inflows. The degree of ethnic tensions, law and order, and government stability are all statistically significant factors affecting net FDI inflows.

Hausmann and Fernández-Arias (2000) analyze the effects of institutional variables in the composition of capital inflows using variables compiled by Kaufmann, Kraay, and Zoido-Lobatón (1999). They find that lack of regulatory quality, government effectiveness and shareholder rights are significant factors explaining reductions in the share of inflows represented by FDI. Using the Institutional Investor Index as a measure of country risk, Raff and Srinivasan (1998) find that in the manufacturing sector there is a -0.55 correlation between country risk and inward FDI. Li and Resnick (2003) find that both property rights protection and democracy-related property rights protection encourage FDI inflows.

In summary, there is concrete evidence from the empirical literature that (1) poor quality of institutions, (2) alternative forms of expropriation, and (3) lack of commitment of policy all have negative effects on FDI inflows. In the next section, we will describe how the theoretical literature attempts to explain these correlations.

3. LESSONS FROM THE THEORETICAL LITERATURE

Most of the theoretical literature assumes that local governments' incentives to expropriate depend on the difference between the benefits of obtaining income from foreign capital (or the ownership of capital) and the opportunity costs of expropriation. Affiliate operation is frequently less successful when managed by the host government rather than by the MNC. This applies specifically to projects in which the hosts import not only physical capital but also foreign entrepreneurship, either in the form of managerial skills or technological know-how.

Under these assumptions, Eaton and Gersovitz (1984) present one of the most influential articles on expropriation theory. They analyze a static economy where competitive investors decide on the amount of foreign investment to be placed in a small open economy. The host country decides whether to expropriate the whole stock of physical capital in order to maximize national income. The cost of such policy is given by the loss in productivity suffered because managerial services are no longer available after expropriation occurs. Foreseeing that their capital might be expropriated ex post, foreign investors will never increase their investment to the level where expropriation becomes optimal. As a result, even though no expropriation occurs in equilibrium, the international allocation of capital is distorted, and FDI remains inefficiently low. Consequently, the ability of the government to expropriate when it lacks the commitment to make binding promises on policy may actually reduce the government's welfare. Empirically, this explains why domestic factor prices may not reflect social returns when the supply of investment is affected by the threat of expropriation. This also supports the finding that commodity trade fails to equate the returns to capital across countries.

Thomas and Worrall (1994) extend this idea to an infinite-horizon economy and characterize the set of self-enforcing agreements between the host government and an MNC (i.e., in a bilateral monopoly environment). The contract specifies the level of investment the MNC should make each period and the amount of output that must be transferred to the host country. The key is that the host government may have a short-term gain by reneging on the contract and expropriating output or capital at any point. In this case, the MNC retaliates by not investing in the future which entails a long-run cost because the domestic economy returns to "autarky." The sustainable contract prescribes that investments should be inefficiently low in the initial periods with no transfers to the host country. Investment rises afterward to a stationary level, in which the host country starts receiving transfers. Investment is pro-cyclical, and transfers are positively serially correlated. Because the temptation to expropriate is larger when output is high, the optimal contract offers more transfers in the future. The back-loading result can be interpreted as a tax holiday, in which the host country exempts investors from tax obligations. It provides some direct transfers and allows for duty-free imports.

Thomas and Worrall's article is closely related to Doyle and van Wijnbergen (1994) who find tax holidays as the outcome of a bargaining game between a foreign investor and a small country, but in which the host country can commit to tax rates for one period. Schnitzer (1999) obtains a similar result by assuming that the foreign investor can switch to production facilities in other countries, rather than assuming commitment to taxes. In contrast to the previous articles, the self-enforcing contract may exhibit overinvestment.

While the previous studies were mostly concerned with explaining the level of expropriation, Aguiar, Amador, and Gopinath (2006) focus on its cyclical properties. The role of the government is to insure the wages of domestic workers, who do not have access to financial markets and are subject to output risk. The government can obtain resources from taxing the MNC's profits (which the authors interpret as an indirect form of expropriation) and redistributing them as lump sum transfers to workers. They show that the combination of lack of commitment and incomplete markets results in policy that generates amplification and prolongation of shocks to output. The government's credibility not to expropriate is scarcest when the economy is in a recession, which depresses investment and prolongs downturns. If the government had the ability to commit to a policy sequence, it would use countercyclical and undistortionary taxes. When it lacks commitment, it distorts foreign investment in bad times and cannot achieve full insurance.

The articles mentioned previously have a common characteristic: governments are benevolent. Policymakers want to maximize welfare (or national output), but they cannot achieve the first best because they are tempted to expropriate too much ex post. The lack of commitment to policy is the main friction in these studies. One aspect that they do not address is that such policies cause redistribution within agents in the host country. Interaction between powerful groups that compete to gain control and appropriate national resources can lead to another source of inefficiencies that distort investment decisions. The political economy game in which a "tragedy of commons" arises, resulting in suboptimal investment levels, is studied in a series of articles by Tornell and co-authors. Tornell and Velasco (1992) explain why, even though poor countries have a higher marginal productivity of capital, they are subject to capital flights toward richer countries. Their main idea is that in countries with weak institutions and poor protection of property rights, some groups can appropriate the returns of other groups by controlling fiscal policy. By investing some of their assets in foreign markets, domestic agents can ensure private access and avoid "overappropriation" (i.e., indirect expropriation) from other groups. Tornell and Lane (1999) use a similar environment to explain how this dynamic interaction between groups leads to a slowdown in economic growth. They show that dilution in the concentration of power ameliorates this problem, a result in contrast to the traditional wisdom in models with a common pool problem. The explanation is based on the fact that groups do not cooperate. So as the number of groups increases, each group must reduce its appropriation rate to make sure its rate of return is no lower than that of its outside option (i.e., investing in the more inefficient informal sector). These articles are closer to ours, due to their emphasis on political factors such as disagreement over redistributive policy across the population of the host country.

Our article is also closely related to Amador (2003), who finds that government borrowing is inefficiently high if there is some probability of losing power in the future. It is also related to Azzimonti (2005), who provides microfoundations in a probabilistic voting model for the shortsightedness of parties in an environment in which the government chooses public investment and the provision of a consumable public good. The underlying force driving the inefficiency of policy is common to all three articles; the difference being that in Azzimonti's environment, investment is chosen by the party in power and taxes are imposed on the domestic group. In the current article, investment is made by foreign investors who have an outside option and the proceeds of expropriating part of it are distributed to a specific group. The article is also related to a body of literature characterizing equilibria that rules out reputation. See, for example, Azzimonti, Sarte, and Soares (2006); Quadrini (2005); or Klein, Krusell, and Ríos-Rull (2004), which characterize Markov-perfect equilibria (the analogous to our equilibrium concept in an infinite-horizon economy). Finally, it is related to a set of political economy models in which redistributive uncertainty results in inefficiencies (see Lizzeri 1999, Alesina and Tabellini 1990, or Battaglini and Coate forthcoming).

4. THE EXPROPRIATION GAME

In this section, we describe the environment and derive our main results. We proceed by specifying the timing and then solving for the subgame-perfect equilibrium through backward induction.

The Environment

The economy is populated by a government, domestic agents, and foreign capitalists. Agents live for two periods. They are endowed with both one unit of time each period and e units of the only consumption good in the economy. We can interpret e as an agent's share of local production (which is not explicitly modeled). Additional output can be produced by identical firms interacting in competitive markets. Shares of these firms are owned by foreign investors who supply capital (denoted by K) but not labor. The opportunity cost of installing capital is given by the world interest rate r^* that could be obtained by investing the funds in riskless bonds in international financial markets. Following Eaton and Gersovitz (1984), we will assume that "managerial services" are the intangible assets that foreign investors bring to the production process: organizational skills, technological knowledge, access to overseas markets, etc. The main difference between managerial skills and physical capital is that the former cannot be expropriated by the government. More importantly, if expropriation occurs, the managerial services of the foreign capitalist are no longer available for production. This implies that any capital expropriated by the government becomes unproductive, because either the domestic worker does not have the necessary skills to run production by himself or because the capital installed by the foreign investor was specific to the manager's skills. Therefore, it cannot be used to produce using the foreign technology.

Production requires two inputs, domestic labor L and capital K and uses the following technology:

Assumption 1 *The production function satisfies*

$$f(K, L) = AK^{\alpha}L^{1-\alpha}.$$

Domestic agents (the *workers*) supply labor inelastically at the competitive wage rate w, and have no international mobility. Each belongs to one of n groups (we can also interpret a group as a collection of individuals residing in one of n districts), with total population normalized to one. Agents are identical, so for symmetry we will assume that there is a measure $\frac{1}{n}$ of agents per group or district. Their preferences over consumption satisfy standard assumptions, as shown below.

Assumption 2 *Instantaneous utility is logarithmic and additively separable, and agents discount the future at rate* $\beta \in (0, 1)$. *Thus,*

$$u(c_1, c_2) = \log(c_1) + \beta \log(c_2).$$

As described in Section 2, expropriation can take two forms: (1) direct expropriation, in which the government takes part or all of the already installed capital, and (2) creeping expropriation, in which transnational corporations are required to pay bribes or licenses that allow them to produce in the host country. Notice that while the former takes place *after* investment decisions have been made, the latter takes place beforehand. This asymmetry will have important implications regarding the effects of electoral uncertainty on expropriation rates.

We will model both forms of expropriation as proportional rates. The government will demand a proportion τ out of total investment to be paid by any firm that intends to produce in the country. Notice that we refer to it as a bribe, but in terms of the modeling technique, it is observationally equivalent to an investment tax. The rate at which installed capital is expropriated ex post will be denoted by θ. Notice that activities are homogeneous in this model, so the host country expropriates all activities at the same rate.[4]

The resources collected by either form of expropriation are used to provide lump sum transfers that can be targeted toward different groups in the population. We will denote the transfer that group i receives, as a function of the expropriation rate, by $T^i(\theta)$.

Assumption 3 *A group's objective, when in power, is to maximize the utility of its members.*

The government expropriates FDI and distributes the proceeds between agents residing in different districts in the country. Two remarks are relevant at this point.

First, even though the expropriation rate by acting as an investment tax distorts the optimal allocation of capital, it serves as an instrument to transfer resources from foreign investors to local workers. The government, who only

[4] We are abstracting from the fact that some sectors are more vulnerable to expropriation than others.

cares about the well-being of domestic agents, might be willing to compromise future production (that will be reduced because inflows of K decrease) in order to collect part of the dividends that would otherwise go to the hands of foreigners. This tradeoff will determine the optimal level of creeping expropriation chosen over time. Notice that the dynamic nature of the game implies that, in general, it would not be optimal for the government to require bribes at a level where investment in the country drops to zero (that is, a τ that drives $I = 0$). Given the assumptions on technology, it would also never be optimal to expropriate capital completely ex post. An important assumption behind this result is the fact that domestic agents cannot produce with the transnational corporation's technology, as described previously.

This environment is a stylized version of an economy where output can be produced with a domestic technology and a (possibly superior) foreign technology. Because we want to focus on the problem of expropriation, rather than on the dynamics of the labor market, we assume that agents are simply endowed with e units of the good and supply labor, inelastically, to foreign firms. It would be interesting to analyze, as an extension, the case in which labor decisions are endogenous and domestic firms compete with transnational corporations for domestic labor. Reallocation of workers from one sector to another after expropriation will cause some distortions—and probably benefit some types of workers while hurting others—that are ignored in the following analysis.

Secondly, since transfers can be targeted toward specific districts, it is reasonable to expect each region to lobby in order to obtain them. The disagreement over how the budget should be allocated across districts can be resolved by some form of voting. One way to model this would be by assuming that there are n parties, each one representing a district that alternates in power according to a Markov process. Amador (2003) presents a model with symmetric parties that want to maximize the group's consumption and face some probability of being in power at each point in time (election dates are uncertain). Once in power, the elected party chooses policy so as to maximize the utility of its constituency. Azzimonti (2005) provides microfoundations for the probabilities in a model of endogenous voting (but in which elections occur at regular intervals). An alternative approach, presented in Battaglini and Coate (forthcoming), assumes that legislators representing a district bargain in congress over redistribution of the budget. These approaches share the property that redistributive uncertainty—captured by the probability of being the decisionmaker in the following period—plays a key role in the level of distortions imposed by policy because of the shortsightedness it introduces.

The sequence of events can be divided into four stages as described below.

Timing

- Period 1:

  1. Creeping Expropriation Stage: the group in power decides the level of bribes they will demand from foreign investors, τ.
  2. Investment Stage: foreign firms decide how much to invest, I, in the host country. Bribes are collected, targeted transfers T_1^i are made, and consumption c_1 takes place.

- Period 2:

  1. Expropriation Stage: one of the groups gains control of the government and expropriates a proportion θ of already installed capital.
  2. Post-Expropriation Stage: the good is produced, wages are paid, targeted transfers T_2^i are made, and consumption c_2 takes place.

Notice that we are assuming that there is no transnational corporation in Period 1, so consumption at that point will be the sum of the endowment an agent possesses and the transfers it obtains from the government (that collected resources in the form of creeping expropriation). We made this assumption to simplify the exposition, but the model can easily be extended to a case in which the government can also expropriate capital installed in the first period of a firm that invested in the country at some point in the past.

We will solve the problem by backward induction, starting from the last stage in Period 2.

The Second Period

Post-Expropriation Stage

This subsection describes the optimization problem faced by the manager of a representative firm. Considering a particular specification for technology and preferences, it characterizes a competitive equilibrium given the expropriation rate and transfers for this economy.

At this stage, the government has already expropriated θK out of the total capital stock, hence the firm produces with the remaining amount of capital $(1-\theta)K \equiv \tilde{K}$. Firms take prices (the wage rate for local workers w) as given, and demand labor in the local market to maximize profits

$$\max f(\tilde{K}, L) + (1-\delta)\tilde{K} - wL,$$

where δ denotes the depreciation rate of capital.

The FOC is

$$f_L(\tilde{K}, L) = w,$$

so labor is paid its marginal productivity. For our given production function, this is equivalent to

$$(1 - \alpha)\, A\tilde{K}^{\alpha} L^{-\alpha} = w.$$

Notice that since $\tilde{K} \leq K$, the wage rate goes down after an expropriation. This occurs because with a lower level of capital installed, workers are less productive (this would hold for any arbitrary function that satisfies $f_{LK} > 0$).

Recall that agents do not have access to capital markets, so their only income is wage income wl, where $l = 1$ is the individual labor supply, plus any transfers T_2^i received from the government. Their budget constraint can be written as

$$c_2^i = e + w + T_2^i.$$

Proposition 1 *A competitive equilibrium given policy* $\{\theta, \{T_2^i(\theta)\}_{i=1}^n\}$*, is a set of prices* $\{w\}$ *and allocations* $\{L, \{c_2^i\}_{i=1}^n\}$ *such that*

1. *consumption of agent* i *satisfies*

$$c_2^i = e + w + T_2^i(\theta),$$

2. *labor supply is* $L = 1$*,*

3. *wages are competitive*

$$w = (1 - \alpha)\, AH^{\eta}\tilde{K}^{\alpha} L^{-\alpha}, \text{ and} \tag{1}$$

4. *the government's budget constraint holds*

$$\sum_{i=1}^{n} \frac{1}{n} T_2^i(\theta) = \theta K.$$

Expropriation Stage

This is the stage in which after a group gains power, it chooses the proportion θ of total capital to be expropriated.[5] A group's objective is to maximize the utility of its supporters. This implies that, while they do not put any weight on the welfare of other regions or groups, policymakers are "benevolent planners" for their own region.[6]

---

[5] Because groups are homogeneous, we can focus on the problem of a representative one.

[6] In the political economy literature, these policymakers are referred to as *partisan*. An alternative approach, also studied in the literature, assumes that the leaders' sole objective is to maximize their probability of controlling the government because they either obtain some egorents from being in power or they can redistribute resources to themselves (*kleptocrats*).

It is assumed that there is no commitment technology: once in power, the group will choose what is best for its constituency from that point on, taking the capital stock as given. This implies that promises made before the political uncertainty is resolved are not credible. In particular, groups cannot credibly promise to transfer resources to other regions in the future. As a result, it is in no group's interest to provide transfers to regions different than its own once it is in power. Mathematically, this implies that group i will optimally set

$$T_2^j = 0 \text{ for } j \neq i.$$

This is the case because (1) groups do not derive utility from the well-being of other regions, and (2) because they cannot sign binding contracts with other groups over policy.

The government balances its budget, so the total amount expropriated is divided among the members of the group controlling the government. In other words,

$$\frac{1}{n} T_2^i = \theta K.$$

The maximization problem of the group in power at this point (where we have omitted the i subscripts for clarity) is

$$\max_{\theta} \; u(c_2) \quad \text{s.t.,}$$

$$\begin{aligned} c_2 &= e + w + T_2, \\ T_2 &= n\theta K, \; and \\ \theta &\leq 1, \end{aligned}$$

where w satisfies equation (1). Replacing the constraints above, we can simplify the objective function to $u(e + w + n\theta K)$. This implies that at the second stage the government maximizes utility by maximizing per capita consumption of the group it represents, so the problem becomes simply

$$\max_{\theta \leq 1} \{e + w + n\theta K\}.$$

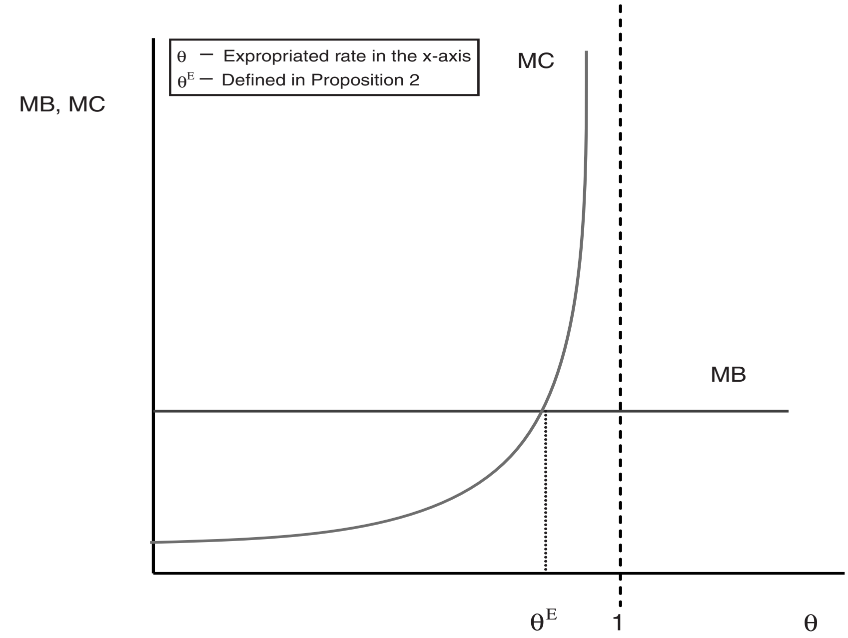
The first-order condition is

$$\frac{\partial w}{\partial \theta} + \frac{\partial T_2}{\partial \theta} \leq 0 \qquad (= 0 \text{ if } \theta < 1).$$

The marginal benefit of increasing the expropriation rate is given by the extra consumption that can be afforded by an increase in the transfer,

$$MB \equiv \frac{\partial T_2}{\partial \theta} = nK > 0.$$

Figure 1 Marginal Benefits (MB) and Costs of Expropriation (MC)



Notice that the marginal benefit is independent of the level of θ. Graphically, it can be represented by a horizontal line (see Figure 1).

The marginal cost is given by a decrease in the agent's labor income due to a reduction in the domestic wage rate,

$$\begin{aligned} MC &\equiv -\frac{\partial w}{\partial \theta} = -\frac{\partial w}{\partial \tilde{K}}\frac{\partial \tilde{K}}{\partial \theta}, \text{ and} \\ &= (1-\alpha)\,\alpha A\,(1-\theta)^{\alpha-1}\,K^{\alpha}. \end{aligned}$$

This function is increasing and convex in the rate of expropriation as long as $\alpha < 1$, as typically assumed with a Cobb-Douglas production technology. Moreover, because the MC becomes infinitely large as $\theta \to 1$, the intersection between the two curves will occur at an interior point (again, refer to Figure 1).

The optimal level of expropriation is found by equating the marginal costs and benefits of increasing θ.

$$(1-\alpha)\,\alpha A\,(1-\theta)^{\alpha-1}\,K^{\alpha} = nK.$$

Proposition 2 *The optimal expropriation rate is given by*

$$\theta^E = 1 - \frac{1}{K}\left[\frac{1}{n}(1-\alpha)\,\alpha A\right]^{\frac{1}{1-\alpha}}.$$

Thus, it is not optimal for any group in power to fully expropriate foreign investment. That is, $\theta^E < 1$.

Since all other groups are identical, the amount of expropriation in this economy is independent of the identity of the group in power. An interesting extension would be to analyze the case in which sectors were heterogeneous, either in their capital intensity or in the ability of the government to expropriate them. In this case, workers would also be heterogeneous and disagree on the rate of expropriation (and not only on where to target the transfers).

The First Period

Recall that there are two relevant stages in this period: investment and creeping expropriation stages. We discuss them in the following section.

Investment Stage

We now move to the decision problem of a foreign firm considering whether the project is worth pursuing in the host country. Expropriation affects this decision on two margins. On the one hand, the cost of investment is increased by the proportion of bribes that will need to be paid to the group in power. On the other hand, the future returns of such investment will be reduced by the fact that some proportion of capital will be expropriated in the second period.

Firms discount the future at the rate $\frac{1}{1+r^*}$, as r^* represents their outside option. The maximization problem faced by an investor at this stage is

$$\max_I \; - I(1+\tau) + \frac{1}{1+r^*}\pi(I) \text{ s.t.,}$$
$$\begin{aligned} \pi(I) &= f(\tilde{I}, L) + (1-\delta)\tilde{I} - wL, \; and \\ \tilde{I} &= (1-\theta)I. \end{aligned}$$

The cost of the investment is incurred today, while the benefits $\pi(I)$ are received next period, which is why they are discounted. The investor knows that for each unit of investment, he will need to pay a proportion τ today in bribes or permits. He also knows that for each unit of capital installed, only a fraction $(1-\theta)$ will be productive: the rest is expropriated by the host country.

The assumption of atomistic competitive investors implies that the action of one of them does not affect the level of expropriation. In other words, each takes θ and τ as given (for the case where the transnational corporation has monopoly, and hence bargaining power, see Doyle and van Wijnbergen 1984 or Thomas and Worrall 1994). The first-order condition for an investor is

$$\underbrace{-(1+\tau)}_{MC} + \underbrace{\frac{1}{1+r^*}\frac{\partial \pi(I)}{\partial I}}_{MB} = 0.$$

Therefore, the transnational corporation equates the marginal cost of investment to the discounted value of the marginal benefit received from its investment opportunity. A marginal increase in installed capital causes an increase in benefits of

$$\frac{\partial \pi(I)}{\partial I} = (1-\theta)\left[f_I(\tilde{I}, L) + (1-\delta)\right].$$

The right-hand side is the marginal increase in production, plus the marginal increase in undepreciated capital, all multiplied by the proportion $(1-\theta)$ that can be utilized. We can interpret $\theta\left[f_I(\tilde{I}, L) + (1-\delta)\right]$ as the opportunity cost of expropriation: it represents the amount of potential benefits that the foreign investor could have obtained if it was not expropriated.

Notice that this value introduces a wedge-distorting investment decision: it produces inefficiencies. This can be seen more clearly from the following equation in which we have replaced the marginal benefit and rearranged the optimality condition:

$$f_I(\tilde{I}, L) + (1-\delta) = \left(1+r^*\right)(1+\tau)\frac{1}{(1-\theta)}.$$

Under our functional forms and noting that in equilibrium $I = K$, the optimality condition becomes

$$\alpha\left((1-\theta)K\right)^{\alpha-1} A + (1-\delta) = \frac{(1+r^*)(1+\tau)}{(1-\theta)}.$$

Lemma 1 *The optimal level of investment under expropriation is*

$$K^E = \left(\frac{\alpha A\,(1-\theta)^{\alpha}}{(1+r^*)\,(1+\tau) - (1-\delta)(1-\theta)}\right)^{\frac{1}{1-\alpha}}.$$

If there was no expropriation, a foreign firm would invest K^{NE} (where NE stands for "no expropriation").

$$K^{NE} = \left(\frac{\alpha A}{(1+r^*)\,(1+\tau) - (1-\delta)}\right)^{\frac{1}{1-\alpha}} > K^E.$$

As expected, expropriation discourages investment in the host country. We can now replace θ by θ^E to find the value of FDI in equilibrium,

$$K^E = \frac{\phi(n)}{1+\tau},$$

where

$$\phi(n) = \frac{1}{1+r^*}\left(A\alpha\frac{(1-\alpha)}{n}\right)^{\frac{1}{1-\alpha}}\left[1-\delta+\frac{n}{1-\alpha}\right]. \tag{2}$$

Creeping Expropriation Stage

At this stage, it is the group in power in Period 1 that decides the level of bribes τ, which it will demand from potential investors. There are two main differences between the tradeoffs faced by policymakers at this point, relative to those faced in the second period, when choosing (ex post) direct expropriation. First, because capital has not yet been installed, FDI is more "elastic." Given the outside opportunities faced by investors, this imposes a constraint on the level of bribes, which in principle, should decrease the temptation to extract too many resources from multinational corporations. Because of this, we would expect creeping expropriation to be less harmful than direct expropriation. On the other hand, the group decides on the level of τ without knowing whether it will be in power next period. This introduces uncertainty over who will have control of the expropriation technology in the second period. More importantly, it introduces uncertainty on the identity of the group receiving the benefits of such expropriation. With probability $1-p$, another group gains control and distributes resources only toward its own region. This second difference with respect to direct expropriation, given by the existence of *redistributional uncertainty*, induces greater expropriation in the present through bribes by *any* group in power in Period 1. Therefore, it is not clear which type of expropriation is more distortive at the end.

Before solving for the optimal level of τ, we need to specify the process by which groups gain control of the government. In this article, we will assume that groups alternate in power according to a stochastic Markov process: the probability of being the decisionmaker next period, given that the group in power today is denoted by p. Notice that this reduced-form specification is silent on whether groups gain control via a democratic process in which parties compete for elections, or the turnover follows from revolutions and coups following a nondemocratic (and possibly violent) process.

Consider the problem faced by a representative group in power in Period 1. It needs to choose the creeping expropriation rate τ on FDI inflows, taking as given the behavior of the domestic sector and foreign firms, as well as competitive prices and aggregates. In particular, it needs to take into account the effects of the bribes and other forms of creeping expropriation chosen based on the following:

1. The consumption of its constituency when the group is in power, because it is maximizing its utility

$$c_1 = e + T_1, \ and$$

$$c_2 = e + w + T_2.$$

2. The consumption of its constituency when the group is *out* of power, because there is a probability that next period a different group is in power,

$$\tilde{c}_2 = e + w.$$

3. FDI inflows, $I = K^E$ (in equilibrium), because foreign investors decide after knowing the level of τ

$$K^E = \frac{\phi(n)}{1+\tau},$$

where $\phi(n)$ is defined in equation (2).

4. Transfers to the region it represents, via the government budget constraint

$$\begin{aligned} T_1 &= \tau n K^E \ \ and \\ T_2 &= \theta^E n K^E. \end{aligned}$$

5. Second period's expropriation rate θ^E,

$$\theta^E = 1 - \frac{1}{K^E}\left[\frac{1}{n}(1-\alpha)\alpha A\right]^{\frac{1}{1-\alpha}}.$$

6. Equilibrium prices, because they affect their constituency's consumption

$$w = (1-\alpha)A[(1-\theta^E)K^E]^{\alpha}.$$

The group solves

$$\max_{\tau} u(c_1) + \beta\{pu(c_2) + (1-p)u(\tilde{c}_2)\},$$

subject to the conditions listed above.

The first-order condition reads as

$$u_c(c_1)\frac{dc_1}{d\tau} + \beta\left\{pu_c(c_2)\frac{dc_2}{d\tau} + (1-p)u_c(\tilde{c}_2)\frac{d\tilde{c}_2}{d\tau}\right\} = 0.$$

When the rate of creeping expropriation increases today, there is a direct effect in agents' consumption—captured in the first term of the expression—since those favored by the group in control receive an increase in the transfer of

$$\frac{dT_1}{d\tau} = \frac{nK^E}{1+\tau}.$$

Firms react to a larger τ by cutting FDI,

$$\frac{dI}{d\tau} = -\frac{K^E}{1+\tau}.$$

This reduces the amount of capital available for production next period (recall that $\frac{dI}{d\tau} = \frac{dK^E}{d\tau}$), which modifies tomorrow's consumption because the next policymaker will face a lower tax base, and thus be forced to reduce the level of transfers T_2. Moreover, from condition 5 this triggers a reduction in direct expropriation (θ^E) as well. The total effect in transfers is given by

$$\begin{aligned} \frac{dT_2}{d\tau} &= n\left[K^E \frac{d\theta^E}{d\tau} + \theta^E \frac{K^E}{d\tau}\right] \textit{ and} \\ &= -\frac{nK^E}{1+\tau}. \end{aligned}$$

Notice that since $(1-\theta^E)K^E$ is independent of τ (see condition 5 above), then from condition 6 so is w. Because second period wages are unaffected by creeping expropriation, the level of consumption when a group is out of power is independent of τ. In other words, $\frac{d\tilde{c}_2}{d\tau} = 0$, so the last term in the first-order condition cancels out. The fact that $\overline{c}_2$ is independent of the level of bribes is a result of the particular assumption on preferences, because under logarithmic utility income and substitution effects cancel out. This results in optimal direct expropriation rates being inversely proportional to the stock of capital, so $(1-\theta)K$ is constant and independent of τ. Replacing u by a logarithmic utility, we obtain the following lemma:

Lemma 2 *Under assumption 2, redistributional uncertainty introduces a wedge in the efficient growth rate of consumption since, in the political equilibrium*

$$c_2 = \beta p c_1.$$

Absent the redistributional uncertainty (i.e., where groups act in a coordinated fashion) the government would choose policy so that $c_2 = \beta c_1$. Because $p < 1$, the equation above shows that the ratio of consumption between the two periods is suboptimally low. In other words, the political uncertainty makes policymakers *too impatient*.

Proposition 3 *Under assumptions 1 and 2, the optimal rate of creeping expropriation is given by*

$$\tau = \frac{\gamma(n) - e\beta p}{(1+\beta p)n\phi(n) - \gamma(n) + e\beta p},$$

where

$$\gamma(n) = n\phi(n)\left[1 + \frac{(1+r^*)(1-\alpha)}{\alpha\left[1-\delta+\frac{n}{1-\alpha}\right]}\right] + e.$$

Expropriation and Political Instability

Political instability refers to the frequency by which groups alternate in power. Countries facing high turnover rates are those where p is relatively small. Why is this the case? Because the probability that any given group remains in control of the government in the second period is low.

In this section, we analyze the implications of political instability on the level of expropriation predicted by the model and contrast it to what the empirical literature has found.

Proposition 4 characterizes how each rate of expropriation changes with p.

Proposition 4 *Under assumptions 1 and 2, we can show that*

1. *creeping expropriation is larger in countries with greater political instability (i.e., low p)*

$$\frac{d\tau}{dp} < 0, \textit{ and}$$

2. *direct expropriation is lower in countries with greater political instability*

$$\frac{d\theta}{dp} > 0.$$

We can understand the intuition behind the negative relationship between the amount of bribes and permits demanded by foreign investors τ, and the probability of keeping control of the government p, by looking at the expression in Lemma 2. When the group in power faces relatively low political instability, the chances of being able to appropriate transfers next period are large. In this case, policymakers want to increase relative consumption (i.e., the ratio $\frac{c_2}{c_1}$). The change in p is equivalent to an increase in the degree of patience of the group in power. Consumption in the second period becomes relatively cheaper, creating a substitution effect toward less consumption today and more consumption tomorrow. Due to market incompleteness, the only way to achieve this transfer of resources is via a reduction in the degree of creeping expropriation today, by lowering T_1 and, thus, c_1. Because transnational corporations bring human capital and technology, they are more efficient in production than the local country. It is then optimal for any group to wait and expropriate after investments have been made when p increases, because the proportion of investment that will not be expropriated ex post will be productive: K^E increases with lower τ rates. If the country had access to borrowing and lending, this effect would be reduced, but nonetheless, present. Therefore, we should expect that countries with low turnover impose relatively low barriers to FDI inflows—that is, require lower bribes and make construction and production permits cheaper to foreign investors.

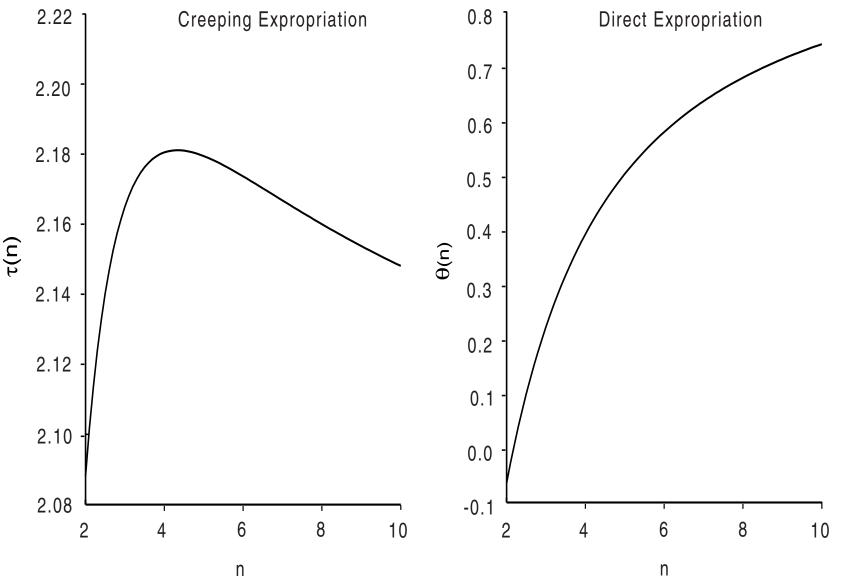
The effect of p on direct expropriation is more subtle and has to do with inter-group manipulation. Because θ is chosen after the political uncertainty has been resolved, it is, in principle, unaffected by p. There is no direct effect of turnover on expropriation ex post. Indirectly, however, increases in p reduce creeping expropriation in the first period and attract more FDI. In other words, K^E increases. Because there is a larger tax base, the MB of expropriating in Period 2 increases. The marginal cost also increases but in a lower proportion (this is due to the Cobb-Douglas technology assumption). As a result, θ goes up. Another way to understand the intuition behind this second result is to consider the costs and benefits of the group deciding today. If p is relatively low, another group will gain control in Period 2 with high probability. If the current group happens to be out of power tomorrow (a likely event), direct expropriation imposes large costs in terms of reduced production and no benefits, because no transfers are received. There are incentives, therefore, to manipulate future decisions by affecting the stock of capital inherited by tomorrow's policymaker and make direct expropriation less attractive. From the expression in condition 5, Section 4, this can be achieved by decreasing K^E. How can the group controlling the government in Period 1 reduce future capital? This can be accomplished by making FDI less attractive—increasing the barriers to its entrance. We should, therefore, expect a negative correlation between political instability and direct expropriation rates.

Notice that this analysis is partial in the sense that we are only considering a once-and-for-all investment decision. There is no action that a government in the second period can take to undo the manipulation of the first period policymaker. In an economy with a longer horizon, in which investment decisions were made every period, the group in power in Period 2 could also demand bribes and permits, and thus break the link between first period bribes and the allocation of foreign capital in the country. That possibility would give groups controlling the government in Period 2 an extra degree of freedom. It would then be interesting to extend the analysis to a case with an infinite-horizon economy.

Expropriation and Concentration of Power

The previous section assumed that differences in political instability only correspond to political factors and were independent of other fundamentals of the economy. In a model where such probability was endogenized, we would expect p to be related to the number of groups in the economy, n. If there were many groups fighting for power, given the same aggregate size of the population, the probability of keeping control of the government would probably be low, and we already know the effects this reduction has on expropriation. On the other hand, a larger value of n implies that if a given group happened

Figure 2 Expropriation and Concentration of Power



to gain control, then the benefits of expropriation *per member* in the group would increase because per capita transfers would be larger. This implies that, in principle, the relation between concentration of power and expropriation could be non-monotonic.

We have calculated how creeping expropriation changes as we reduce the concentration of power for a numerical example (the parameter values were not calibrated but rather chosen to illustrate our point). The probability of staying in power faced by any group is assumed to satisfy

$$p = \frac{1}{n} + \xi,$$

where ξ represents an "incumbency advantage" term, reflecting the fact that the group in power has greater chances to gain control next period than any other group in the opposition. The political economy literature has documented the existence of such an advantage in democratic elections. In more authoritarian systems, we often see groups or families in control of the government for long periods of time because they have access to military force and other means of repression. Increases in ξ can be interpreted as changes in political instability not related to the concentration of power, which were studied in the previous section, whereas the effects of concentration can be analyzed separately by looking at the effects of changes in n.

Inspection of Figure 2 tells us that when there is relatively large concentration of power (i.e., n close to 2), increases in the number of groups result in more expropriation of both types. This happens because larger values of n reduce p and, from the intuition in the previous section, this encourages creeping expropriation activities, that is, rises in τ. On the other hand, when there is little concentration of power, increases in the number of groups in a given economy result in *lower* levels of τ. While the probability of remaining in power decreases with n, transfers per capita increase, but in a larger proportion and, therefore, dominate. Even though each group is less likely to stay in power, the benefits of expropriating in Period 1 more than compensate the costs driven by an increase in the risk of losing control of the government in Period 2. This result is different from the one found in the previous section, and it gives a direct testable implication of the model. If countries have greater political instability because there is low incumbency advantage, more creeping expropriation is to be expected. If, on the other hand, it is due to the composition of competing groups, and there is a relatively large number of them, then we should expect *less* creeping expropriation as political instability increases.

5. CONCLUSIONS

We reviewed the empirical evidence on the effects of expropriation on FDI inflows, mainly focusing on developing countries. We then discussed theoretical models explaining how the quality of institutions affects FDI and growth. In particular, we described how the different frictions present in the political process result in policies that discourage FDI inflows. Finally, we presented a simple model that sheds some light on the effects of expropriation on FDI under: (1) lack of commitment to policy from the government, (2) redistributional uncertainty resulting from stochastic alternation of groups in power, and (3) the interaction between alternative forms of expropriation. The main contribution of this work is twofold: the analysis of a model in which both direct and indirect forms of expropriation are present and the study of how the two types of expropriation relate to political instability. We also discussed the effects of the concentration of power on the incentives to use each type of expropriation and their resulting effects on investment.

REFERENCES

Aguiar, Mark, Manuel Amador, and Gita Gopinath. 2006. "Efficient Expropriation: Sustainable Fiscal Policy in a Small Open Economy." Federal Reserve Bank of Boston, Working Paper 06–9.

Alesina, Alberto, and Guido Tabellini. 1990. "A Positive Theory of Fiscal Deficits and Government Debt." *Review of Economic Studies* 57 (3): 403–14.

Amador, Manuel. 2003. "A Political Economy Model of Sovereign Debt Repayment." Mimeo, Stanford Graduate School of Business.

Asiedu, Elizabeth. 2006. "Foreign Direct Investment in Africa: The Role of Natural Resources, Market Size, Government Policy, Institutions and Political Instability." *World Economy* 29 (1): 63–77.

Azzimonti, Marina. 2005. "On the Dynamic Inefficiency of Governments." Mimeo, University of Iowa.

Azzimonti, Marina, Pierre-Daniel G. Sarte, and Jorge Soares. 2006. "Distortionary Taxes and Public Investment When Government Promises Are Not Enforceable." Mimeo, Federal Reserve Bank of Richmond.

Battaglini, Marco, and Steve Coate. Forthcoming. "A Dynamic Model of Legislative Bargaining." *The American Economic Review*.

Biais, Bruno, and Enrico Perotti. 2002. "Machiavellian Privatization." *The American Economic Review* 92 (1): 240–58.

Busse, Matthias, and Carsten Hefeker. Forthcoming. "Political Risk, Institutions and Foreign Direct Investment." *European Journal of Political Economy*.

Chifor, George. 2002. "Caveat Emptor: Developing International Disciplines for Deterring Third Party Investment in Unlawfully Expropriated Property." *Law and Policy in International Business* 33 (2).

Corkran, Douglas E. 1991. "Tanzania Opens its Doors to U.S. Business." *Business America* 1 (July): 11.

Daude, Christian, and Ernesto Stein. 2001. "The Quality of Institutions and Foreign Direct Investment." Paper presented at the 2001 Annual Meeting of the Inter-American Development Bank. Santiago, Chile.

Doyle, Christopher, and Sweder van Wijnbergen. 1994. "Taxation of Foreign Multinationals: A Sequential Bargaining Approach to Tax Holidays." *International Tax and Public Finance* 1 (3): 211–25.

Eaton, Jonathan, and Mark Gersovitz. 1984. "A Theory of Expropriation and Deviations from Perfect Capital Mobility." *Economic Journal* 94 (373): 16–40.

Gordon, H. Scott. 1954. "The Economic Theory of a Common-Property Resource: The Fishery." *Journal of Political Economy* 62 (2): 124–42.

Hausmann, Ricardo, and Eduardo Fernández-Arias. 2000. "Foreign Direct Investment: Good Cholesterol?" Inter-American Development Bank, Research Department Working Paper No. 417, Washington, DC.

Hines, James R. 1995. "Forbidden Payment: Foreign Bribery and American Business After 1977." NBER Working Paper No.W5266.

Janeba, Eckhard. 2002. "Attracting FDI in a Politically Risky World." *International Economic Review* 43 (4): 1,127–55.

Jensen, Nathan. 2005. "Measuring Risk: Political Risk Insurance Premiums and Domestic Political Institutions." Mimeo, Washington University.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. "Governance Matters." World Bank. Mimeo WPS2196.

Kindleberger, Charles P. 1969. *American Business Abroad: Six Lectures on Direct Investment*. New Haven, CT: Yale University Press.

Klein, Paul, Per Krusell, and Jose-Victor Ríos-Rull. 2004. "Time-Consistent Public Expenditures." CEPR Discussion Paper No. 4582.

Kobrin, Stephen J. 1980. "Foreign Enterprise and Forced Divestment in LDCs." *International Organization* 34 (1): 65–88.

Kobrin, Stephen J. 1984. "Expropriation as an Attempt to Control Foreign Firms in LDCs: Trends from 1960 to 1979." *International Studies Quarterly* 28 (3): 329–48.

Lancaster, Kelvin. 1973. "The Dynamic Inefficiency of Capitalism." *Journal of Political Economy* 81 (5): 1,092–109.

Li, Quan. 2004. "Democracy, Autocracy, and Tax Incentives to Foreign Direct Investors: A Cross-National Analysis." Mimeo, Pennsylvania State University.

Li, Quan, and Adam Resnick. 2003. "Reversal of Fortunes: Democratic Institutions and Foreign Direct Investment Inflows to Developing Countries." *International Organization* 57 (1): 175–212.

Lizzeri, Alessandro. 1999. "Budget Deficits and Redistributive Politics." *Review of Economic Studies* 66 (4): 909–28.

Lucas, Robert E., Jr. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *The American Economic Review* 80 (2): 92–6.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110 (3): 681–712.

Minor, Michael S. 1994. "The Demise of Expropriation as an Instrument of LDC Policy, 1980-1992." *Journal of International Business Studies* 25 (1): 177–88.

Moran, Theodore H., ed. 2004. *International Political Risk Management: The Brave New World.* Washington, DC: The World Bank.

Organization for Economic Cooperation and Development. 2004. "Indirect Expropriation and the Right to Regulate in International Investment Law." OECD Working Paper on International Investment No. 2004/4.

Quadrini, Vincenzo. 2005. "Policy Commitment and the Welfare Gains from Capital Market Liberalization." *European Economic Review* 49 (8): 1,927–51.

Raff, Horst, and Krishna Srinivasan. 1998. "Tax Incentives for Import-Substituting Foreign Investment: Does Signaling Play a Role?" *Journal of Public Economics* 67 (2): 167–93.

Robertson, Dario, and Xiaokang Chen. 1990. "New Amendments to China's Equity Joint Venture Law." *East Asian Excecutive Reports* 15 (April): 9–13.

Schlemmer-Schulte, Sabine. 1999. "The World Bank Guidelines on the Treatment of Foreign Direct Investment." In *Legal Aspects of Foreign Direct Investment,* eds. Daniel D. Bradlow and Alfred D. Escher. The Hague, The Netherlands: Kluwer Law International: 87, 99–100.

Schnitzer, Monika. 1999. "Expropriation and Control Rights: A Dynamic Model of Foreign Direct Investment." *International Journal of Industrial Organization* 17 (8): 1,113–37.

Shleifer, Andrei, and Robert W. Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108 (3): 599–617.

Thomas, Jonathan, and Timothy Worrall. 1994. "Foreign Direct Investment and the Risk of Expropriation." *Review of Economic Studies* 61 (1): 81–108.

Tornell, Aarón, and Andrés Velasco. 1992. "The Tragedy of the Commons and Economic Growth: Why Does Capital Flow From Poor to Rich Countries?" *Journal of Political Economy* 100 (6): 1,208–31.

Tornell, Aarón, and Philip R. Lane. 1999. "Voracity and Growth." *The American Economic Review* 89 (1): 22–46.

United Nations. 1996. *World Investment Report*. New York and Geneva: The United Nations.

Wei, Shang-Jin. 2000. "How Taxing is Corruption on International Investors?" *The Review of Economics and Statistics* 82 (1): 1–11.