Central Banking in a Democracy

Alan S. Blinder

On September 26, 1996, Alan S. Blinder presented the following speech at the Federal Reserve Bank of Richmond. The text of the speech is printed below after J. Alfred Broaddus's introduction of the speaker.

It is a pleasure to welcome all of you this afternoon for this latest program in our series of occasional lectures by distinguished economists on major economic policy issues. Largely for convenience, we hold these programs here at the Richmond Fed, and we are delighted to host them. But let me remind you that they are jointly planned and funded with the three university business schools here in Richmond. I'd like to introduce my colleagues in this endeavor: Dr. Al Altimus, Dean of the Lewis School of Business at Virginia Union University; Dr. Randolph New, Dean of the Robins School of Business at the University of Richmond; and Dr. Howard Tuckman, Dean of the School of Business at Virginia Commonwealth University. It's been a great pleasure working with these folks over the years and I think it's been a very productive collaboration. I would also like to recognize my colleague, Marvin Goodfriend, who is Senior Vice President and Director of Research here at the bank. Marvin is my principal adviser and has played a leading role in planning and putting on these programs in recent years.

It's a particular personal pleasure and an honor to introduce our speaker. Alan Blinder, to put it bluntly but accurately, is one of the most distinguished macroeconomists in the world today. He has done about everything any economist and even a leading economist would do. He earned a Ph.D. from a leading economics department, MIT. He has taught and is teaching at a top university, Princeton, which also happens to be his undergraduate alma mater. He has published numerous scholarly articles in professional journals, a leading

The views expressed are those of the speaker and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

economics textbook, and a number of other books, including a little gem on major economic policy issues called *Hard Heads, Soft Hearts.* Finally, he has served as an economic policymaker at the highest level. Early in his career, in the mid-'70s, he was briefly Deputy Assistant Director of the Congressional Budget Office when it was just getting started. More recently, of course, he was a member of the President's Council of Economic Advisers in the first year and a half of the Clinton Administration. In that capacity, he was a leading economic policy adviser to the Administration. Subsequently he was our colleague at the Fed, serving as Vice Chairman of the Board of Governors of the Federal Reserve System from June 1994 until he departed early this year to return to Princeton, where he is the Gordon S. Rentschler Memorial Professor of Economics. Some of you who have been attending these lectures regularly will remember that we had Alice Rivlin here not too long ago. She has now become the Vice Chairman of the Board of Governors. We wanted to give an equal opportunity to *former* Vice Chairmen of the Federal Reserve Board.

It's not an idle compliment when I say that as a veteran Fed employee, I greatly enjoyed Alan's all-too-brief tenure at the Fed. Let me share with you a little Fed secret. It's not about interest rates, but about the Federal Open Market Committee, which as you know is the main policymaking body in the Fed. It's a strong and I believe effective committee, but it is not always a terribly lively committee. It was, however, a much livelier committee when Alan was part of it. He challenged us—and also helped us—to confront issues objectively with careful and solid economic analysis. He helped take the edge off the debates that we had in the committee during his tenure with keen and well-timed humor. And he raised the level of discussion in the committee during his tenure. I miss his input very much. It's good to have him back in Richmond. He was here about a year-and-a-half ago and gave a great lecture to the Virginia Bond Club. Please join me in welcoming back Dr. Alan Blinder.

Al, thank you very much for that fine introduction. I want to talk this afternoon about the role of the Federal Reserve in society in very broad terms and, along the way, to make a few rather more specific points. Then I will be glad to entertain questions about what interest rates will do next week, a subject about which I know nothing! But neither does anyone else, so we are all on an equal footing on the subject.

WHO DOES THE FED SERVE?

Relative to their economic and therefore social importance, central banks must be among the least well understood institutions in the entire world. For example, I have been told that millions of Americans still think that the Federal Reserve

System is a system of government-owned forests and wildlife preserves where, presumably, bulls and bears and hawks and doves frolic together in blissful harmony. Having spent 19 months there, I can assure you that that is not the case. The Federal Reserve is an institution that touches almost everyone in America, plus many people outside America, but is itself touched or even seen by relatively few. But its traces are everywhere. Every time you pay or receive paper currency, you are using a “Federal Reserve Note”—a debt obligation of the Fed. Unbeknownst to most of you, your checks are also probably cleared through a regional Federal Reserve Bank; and, if you bank here in Richmond, through this regional Federal Reserve Bank.

When you read in a newspaper ad that a certain bank will pay you 5.7 percent on a certificate of deposit or you hear on television that an automobile company this week is offering 4.9 percent financing, you are seeing tangible evidence of the Fed’s regulatory hand at work. Very few people have any idea that it is the Fed that tells banks and auto finance companies how to calculate and advertise those numbers. And even fewer know that the Fed doesn’t always get it right!

The interest rates themselves, while set in free markets, are heavily influenced by the Fed’s monetary policy. Most Americans these days know that, but few can tell you how that black magic is performed. Even fewer people understand how the Fed’s interest rate decisions impact on the overall economy and therefore influence how many people will find jobs, how many will be laid off, how many businesses will succeed, how many will fail. Most economists will attest to the fact that the Fed has far more influence over these matters than the President and Congress.

The Federal Reserve System has a governance structure that is at least odd and perhaps even byzantine. While most countries in the world have one central bank, we have 12—the one here in Richmond and 11 others. These regional Federal Reserve Banks are, in a legal sense, private corporations. They have presidents, in this case Al Broaddus, and boards of directors. They even have shareholders. And while these corporations are extremely profitable—to the tune of over $20 billion a year for the 12 of them together—their shareholders do not reap the benefits. Instead, the Fed’s prodigious profits are turned over to the United States Treasury—a very friendly gesture. Atop this organization of 12 putatively private corporations, there sits a seven-member Board of Governors in Washington, whose members are not elected by any of the stockholders, but rather are politically appointed. A very, very curious organizational structure.

So the question arises: Who does the Fed serve? Congress and the President? Most certainly not. Although the Fed is a creature of Congress, and its governors are all presidential appointees, the Fed does not exist to do their bidding. After all, that would make a mockery of the doctrine of central bank independence.

What about the banks? Well to some extent, the answer must be yes. The Fed is a bank for banks. It sells to these banks a variety of services, many of them in direct competition with private suppliers. The Fed is also deeply concerned with the health of the banking and payments system and will, when necessary, take strong steps to safeguard it. However, the Federal Reserve is also the supervisor of thousands of banks, either directly or indirectly, through their bank holding companies. (The Fed supervises all the bank holding companies.) It is a very odd arrangement when you think about it: The Fed is regulating its own customers. There are a lot of businesses in America that would like to regulate their own customers, but very few get to do that. In my view, it is a great mistake for the Federal Reserve to see itself as a service organization for the benefit of banks, however. It is a mistake that people in the Federal Reserve System make occasionally, but fortunately not very often.

Does the Fed serve the financial markets? As the nation's central bank, the Fed is naturally and certainly the ultimate guardian and protector of the entire financial system. In times of acute market distress, the Fed stands ready to play its classic role as lender of last resort. In more normal times, the Fed worries about such things as the integrity of the markets, financial fragility, speculative bubbles, the value of a dollar, and a host of other things. As I used to say when I was Vice Chairman of the Fed, we get paid, though not very much, to worry about *everything*.

But, in my view, none of these choices—not the President, not the Congress, not banks, not the financial markets—adequately describes the Fed's true constituency. In my view, that constituency can only be the entire nation. While I was on the Federal Reserve Board, I often said that I viewed myself as working for 260 million Americans. Given the central bank's broad reach and pervasive influence, no narrower constituency seemed appropriate. So I want to talk this afternoon about what the Federal Reserve does and should do to serve the national interest.

I think it would surprise most of you to learn that a time-and-motion study of the daily lives of Federal Reserve Governors would reveal that most of their efforts are devoted to bank regulatory issues, broadly defined. Most of this business is routine, extremely familiar, and intensely interesting to the banking industry, and totally unknown, deeply obscure, and generally quite boring to everybody else in society. This is the Federal Reserve that nobody knows. So as not to bore you with these matters, I will skip directly to the Federal Reserve that everybody knows, for nowhere is the Federal Reserve's public service role more visible than in the conduct of monetary policy. It is monetary policy that puts the Fed in the news constantly, and occasionally puts it in the middle of a political maelstrom.

If you don't live your life in the financial world, it is almost impossible to imagine how tightly focused the media and the markets are on the Federal Reserve. Fixated is not too strong a word. "Federal Reserve fetish" has not yet come into current use as a term of art, but I think it describes a lot of the behavior of the financial press and people in the financial markets. To the financial press, a Federal Reserve Governor is more engaging than a movie star. (Think about that one for a while!) When I was Vice Chairman of the Federal Reserve Board, I simply came to expect to find 15, 20, or 25 reporters, plus several TV cameras, waiting any time I made a public appearance—no matter how boring my speech was going to be. (This, as you've noticed, doesn't happen to me anymore.)

Things were not always this way. The story is told that the only way President Kennedy could remember the difference between *monetary* policy and *fiscal* policy was that the letter "M" for monetary was also the first letter of the name of the Fed Chairman at that time, William McChesney Martin. Times have sure changed. I can assure you that President Clinton had no such problem, and neither did President Bush.

THE GOALS OF MONETARY POLICY

Just how is the Fed supposed to serve the national interest with this strange instrument called monetary policy? Under the terms of the Federal Reserve Act, as amended, Congress has directed the Fed to promote "maximum employment, stable prices, and moderate long-term interest rates." That sounds like three goals, but the phrase is often called the Fed's "dual mandate" because the interest rate objective is considered redundant. Price stability will almost certainly bring low long-term interest rates in its wake.

At this point, I need your indulgence for a very brief Economics 101 lecture on how monetary policy affects employment and inflation. It all works roughly as follows.

In the short run, employment is largely determined by total spending in the economy. Interest rates are one, though not the only, important determinant of that spending. So the Federal Reserve, via its effect on market rates of interest, exerts considerable indirect influence over employment and unemployment. But the process takes time. As economists put it, monetary policy works with long lags. While the lagged effects of monetary policy on unemployment are distributed through time—a little now, a little more the next quarter, and so on—it won't hurt you to think of them as taking about a year or two.

Changes in inflation, up or down, are largely determined by the balance between total spending, which is heavily influenced by monetary policy, and the economy's capacity to produce, which is not. If spending falls short of the economy's productive capacity, as happens in a recession, inflation will fall. If spending overshoots capacity, as sometimes happens in a boom, inflation will rise. But the lag from monetary policy decisions to inflation is even longer than the lag from monetary policy to employment because monetary policy first has to affect spending, and then spending must affect inflation. Think of the whole process—from a decision of the Federal Reserve on monetary policy to the reaction of inflation—as taking more than two years.

The central dilemma of monetary policy is this: Unless inflation is below the Federal Reserve's long-run target, which hasn't been true in a very long time, there is a short-run trade-off between the two goals—maximum employment and stable prices—that are set forth in the Federal Reserve Act. To push inflation lower, the Fed must make interest rates high enough to hold total spending below the economy's capacity to produce. But if it does that, the Federal Reserve will be reducing employment, contrary to the dictum to pursue "maximum employment." So monetary policy is forced to strike a delicate balance between the two goals. It is an excruciatingly difficult decision, with a great deal at stake. As a former holder of my former office once quipped, "That's why they pay them the big bucks!"

THE TRADE-OFF AND JACKSON HOLE

Early in my term as Vice Chairman of the Fed, I allegedly stirred up a controversy at a Federal Reserve conference in Jackson Hole, Wyoming, by acknowledging this trade-off explicitly. The context is important to an understanding of what happened, because the subject of that conference was *reducing unemployment*. Being a central banker at the time, I thought it was appropriate for me to address the role of central banks in that task. In my brief remarks, I noted that monetary policy actions have a profound effect on employment. I also suggested that a central bank could do its part to achieve low unemployment by pushing the nation's total spending up to the level of capacity, but not further. I observed that the Fed's dual mandate could reasonably be interpreted in precisely that way. So I endorsed that mandate as eminently reasonable instructions for the Congress to have given the Fed, rejecting the alternative of concentrating exclusively on price stability and ignoring unemployment.

Nothing I said at Jackson Hole that day was really controversial, and certainly nothing was original. My conceptualizations of monetary policy's role and of the trade-off between inflation and unemployment were totally conventional. My endorsement of the Fed's dual mandate meant that the Vice Chairman of the Federal Reserve was publicly endorsing the Federal Reserve Act. Now there's news for you! Furthermore, my implied "advice" to central bankers was fully consistent with the practices of central banks all over the world, regardless of what they preach. Indeed, I think a very fair academic critique of my little talk that day would have labeled it banal. Had a student in a course submitted that talk to me as a paper I think I would have said, "There is not an original idea here. You have to be able to do better than this."

About a dozen financial journalists were in the audience that day, and all but one of them heard it that way. But Keith Bradsher of *The New York Times* decided that he had just heard "a big story." I had, he was led to believe by some anonymous whispers, violated the sacred trust of central bankers by saying a few obvious things out loud. He told readers of *The Times* that I had publicly clashed with the Fed Chairman. That's funny; Alan Greenspan was sitting right there as I spoke, and he didn't hear it that way at all. I know, because the two of us had breakfast together the next morning, and he never indicated that I had said anything unusual—which I hadn't. No matter. On a slow news day in August Bradsher's story from Jackson Hole wound up on page one of *The New York Times*.

Media firestorms have a life of their own and, until you have been the subject of one, it is hard to imagine what they are like. For more than a month, a seemingly unending barrage of stories appeared in newspapers, magazines, over the financial wires, and even on TV and radio. I was made "controversial," which is one of the ways they try to stick the knife in you in Washington. The Fed, the public was told, had an outspoken new Vice Chairman who had broken several central banking taboos and publicly tangled with his Chairman. The hysteria reached a crescendo with truly a malicious attack published in both *Newsweek* and *The Washington Post* by Robert Samuelson, who decided—without ever bothering to call me up even once to talk about my views—that I was unfit both morally and intellectually to lead the Fed. One would be okay, but *both* morally and intellectually? Whoever said that serving in the government isn't fun?

I recount this episode not to dredge up the ghosts of irresponsible journalists' pasts, but for three reasons that are closely related to today's topic. The first is to give you a little window into what can happen when a Federal Reserve Governor publicly endorses the view that the Federal Reserve should be serving the national interest rather than just the parochial interests of the bond market. But I must insist that serving the national interest is the *only* correct way to conceptualize the Fed's mission; to me, the issue is not open to either compromise or debate.

The second reason is to tell you that I remain totally unrepentant and never retreated one inch from the position that I enunciated that day—not in public, and not inside the Federal Reserve. What I said that day was true then, and it is true now. There is abundant evidence that Keynes was right back in the '30s when he said that modern industrial economies are not sufficiently self-regulating. They need a little help. Total spending sometimes roars ahead of productive capacity, which leads to accelerating inflation. And total spending sometimes lags behind productive capacity, leading to unemployment.

In principle, either fiscal policy—the government's taxing and spending policy—or monetary policy could serve as the balance wheel, propping up demand when it would otherwise sag and restraining it when it threatens to race ahead too rapidly. In practice, however, monetary policy is the only game in town nowadays. And when I say "in town," I don't mean just in Richmond or just in the United States—I mean all over the industrial world. The reason is the same here and in Europe: The need to reduce large fiscal deficits dictates that budget policy remain a drag on total spending for the foreseeable future, regardless of the state of the macroeconomy. With the fiscal arm of stabilization policy thereby paralyzed, a central bank that decides to concentrate exclusively on price stability is, in effect, throwing in the towel on unemployment.

So, to me, the argument for the Fed's dual mandate is both straightforward and convincing. The central bank exists to serve society. The public cares deeply about fluctuations in the pace of economic activity. And well-executed monetary policy has the power to mitigate fluctuations in employment. As the mathematicians say, "QED." Fortunately, almost all central bankers accept this argument nowadays, notwithstanding a great deal of misleading rhetoric to the contrary.

That leads me straight to the third reason for telling you the Jackson Hole story. As a citizen of a democracy, I have always found it intolerable for the government to deceive the governed. As a public servant, I also found it unconscionable. And I see no reason whatsoever why the central bank should have a special exemption from the requirement to level with the public.

CREDIBILITY

It is sometimes argued, to the contrary, that honest acknowledgment of the trade-off between unemployment and inflation, and of the central bank's concern with each, would rattle the financial markets—which want to believe that the central bank cares only about low inflation. This argument is nonsense. Both market participants and the financial press know the score and are far too sophisticated to be taken in by ritualistic rhetoric. I remember very well a conversation I had with a very smart financial reporter shortly after I left the Fed. He said that he has learned over the years to ignore what the Fed *says* and watch what it *does*. I had to concede that he was right, but it troubled me a great deal that the two would be so different. In my view, they should be a matched pair.

There is much talk at the Federal Reserve, as in the central banks all over the world, about the importance of *credibility*, which, according to the dictionary in my office, is "the ability to have one's statements accepted as factual or one's professed motives accepted as the true ones." Let me read the last phrase again: "one's professed motives accepted as the true ones." Precisely the point! Why is credibility considered so important?

The main reason, in my view, is that a central bank is a repository of enormous power over the economy. And if the central bank is independent, as the Federal Reserve is, this power is virtually unchecked. Such power is a public trust, assigned to the bank by the body politic through its elected representatives. In return, the citizens and their elected representatives have a right to expect—indeed to demand—that the bank's actions match its words. And matching deeds to words is, to me and to my dictionary, the hallmark of credibility.

CENTRAL BANK INDEPENDENCE

The Fed's role as the macroeconomic balance wheel is terribly important because it palpably effects people's lives. Stabilization policy is not something abstract; it is about how many jobs there will be, how many businesses will succeed. In my view, it is far and away the most important thing a central bank does for or to its society. And I felt that responsibility keenly every day that I served as Vice Chairman of the Fed, as I know Al Broaddus still does in his role as a member of the Federal Open Market Committee. Society, therefore, has a strong interest in seeing to it that the central bank does its job well. Evidence collected in recent years suggests that making the central bank more independent should help.

Before elaborating on this point, however, I need to define what I mean by an independent central bank—because there is no agreed-upon definition. To me, the term connotes two things.

The first is that the central bank is free to decide *how* to pursue its goals. This freedom does not mean that the Bank gets to select the goals on its own. On the contrary, in a democracy it seems not just appropriate, but virtually obligatory, that the political authorities should set the goals and then instruct—and I use that verb advisedly—the central bank to pursue them. If it is to be independent, the bank must have a great deal of discretion over how to use its instruments in pursuit of its assigned objectives. But it does not have to have the authority to set the goals by itself. Indeed, I would argue that giving the bank such authority would be an excessive grant of power to a bunch of unelected technocrats. In a democracy, the elected representatives of the people should make decisions like that. The central bank should then serve the public will.

The second critical aspect of independence, in my view, is that the central bank's decisions cannot be countermanded by any other branch of government, except under extreme circumstances. In our system of government, neither the President nor the Supreme Court can reverse a decision of the Federal Open Market Committee. Congress can, in principle, reverse such a decision, but only if it passes a law that the President will sign (or by overriding a presidential veto). This makes the Fed's decisions, for all practical purposes, immune from reversal; and, indeed, they never have been reversed. Without that immunity, the Fed would not really be independent, for its decisions would stand only as long as they did not displease someone more powerful.

In recent years, considerable empirical evidence has accumulated in support of the idea that macroeconomic performance is superior in countries that have more independent central banks. Researchers here and in other countries have developed several creative ways to measure central bank independence. Such measures include the bank's legal status, the rate of turnover of its leaders, the legal mandate in the bank's charter (for example, whether it is directed to pursue price stability), and answers to a questionnaire about its organizational structure. The clear weight of this evidence, and by now there is a lot of it, is that countries with more independent central banks have enjoyed lower average inflation without suffering lower average growth. This finding is, of course, completely consistent with economists' general view that, while there is a short-run trade-off between inflation and unemployment, there is no long-run trade-off.

These research results on the benefits of central bank independence raise a provocative question: Why is it that central banks possessing greater independence produce superior macroeconomic results on average? I want to suggest three reasons, all closely related.

First, as I emphasized in my brief Economics 101 lecture a couple of minutes ago, the effects of monetary policy come with long lags. So, to conduct monetary policy well, you must look far in the future and then wait patiently for the results. Farsightedness and patience, I dare say, are not the strong suits of the political process in a democracy. But they are absolutely essential to pursuing a successful monetary policy.

Second, and related to the time-horizon question, inflation-fighting has the characteristic cost-benefit profile of a long-term investment: You pay the costs of disinflation up front, and you reap the benefits—lower inflation—only gradually through time. So, if politicians were to make monetary policy on a day-to-day basis, they would be sorely tempted to reach for short-term gains at the expense of the future—that is, to inflate too much. Aware of this temptation, many governments wisely depoliticize monetary policy by delegating authority to unelected technocrats with long terms of office, thick insulation from the hurly-burly of politics, and explicit instructions to fight inflation.

Third, and related to this point about technocracy, the conduct of monetary policy is at least somewhat technical. It is a bit like shooting a rocket to the moon, though not nearly as exact. Very few elected officials in this or other countries have much understanding of how the monetary transmission mechanism works, of the long lags that I have mentioned, or of a variety of other technical details about monetary policy. So countries can probably get higher-quality monetary policy by turning the task over to trained technicians, subject, of course—and this is important in my view—to political oversight.

CENTRAL BANK INDEPENDENCE AND DEMOCRACY

At this point, a very deep philosophical question arises: Isn't all this profoundly undemocratic? Doesn't assigning so much power to unelected technocrats contradict some fundamental tenets of democratic theory? It is a legitimate question. My answer is: If you assign this power well, it needn't be antidemocratic. And I want to conclude this lecture with a detailed defense of that answer. The question is: How can an independent central bank be rationalized within the context of democratic government? My recipe comes in six parts.

First, we all know that, even in democracies, certain decisions are reserved to what is sometimes called the "constitutional stage" of government, rather than left to the daily legislative struggle. These are basic decisions that we do not want to revisit often; they should, therefore, be hard to reverse. So, for example, amending the U.S. Constitution requires much more than majority votes of both houses of Congress. The Founding Fathers thereby made it almost, but not quite, impossible to change certain basic provisions of law. And they meant it that way; it wasn't an accident.

Similarly with monetary policy. The Fed's independence, which derives from authority delegated by Congress, makes it very difficult, but not quite impossible, for elected officials to overrule or influence a monetary policy decision. Wise politicians made a once-and-for-all decision years ago to limit their own power in this way just as, for example, the Constitution made it very difficult to change the length of the President's term of office. The reasoning was precisely the same as that which led Ulysses to tie himself to the mast. He knew he would get better long-run results even though he wouldn't feel so good about it in the short run!

The second ingredient that helps make central bank independence consistent with democratic theory is something I emphasized earlier: The bank's basic goals are chosen by elected politicians, not by unelected technocrats. So, for example, when people suggest to me that the Fed should content itself with 3 percent inflation, I always answer, "the Federal Reserve Act, which is the law of the land, says 'stable prices,' not 'pretty low inflation.' " If the citizens think that is wrong, they should get the law changed. Until they do, the Federal Reserve should obey the law.

Third, the public has a right to demand honesty from its central bankers. This, again, is a point I made earlier in discussing the idea of credibility, which I defined as matching deeds to words. The central bank, in my view, owes this to the body politic in return for the broad grant of power it enjoys.

The fourth ingredient is closely related to this last point. I call it *accountability*, or perhaps just *openness*. Monetary policy actions have profound effects on the lives of ordinary people. In my view, a central bank in a democracy therefore owes these folks an explanation of what it is doing, why it is doing it, and what it expects to accomplish by its actions. As I often said while I was at the Fed, "It's their economy, not ours." By offering a reasonably full and coherent explanation of its actions, the central bank can remove much of the mystery that now surrounds monetary policy, enable interested parties to appraise its decisions contemporaneously, and then—importantly—allow outsiders to judge its success or failure after the fact, for the verdict of history is the only one that ultimately matters.

Let me assure you that greater openness is not a popular cause in central banking circles, where some see mystery as essential to effective monetary policymaking. Making the central bank more open and accountable, it is alleged, may subject it to unwelcome scrutiny that could threaten its independence.

I couldn't disagree with this argument more. In fact, I think it gets matters exactly backward. To me, public accountability is a moral corollary of central bank independence. In a democratic society, the central bank's freedom to act implies an obligation to explain itself to the public. Thus independence and accountability are symbiotic, not conflicting. Accountability legitimizes independence within a democratic political structure.

Nor, by the way, do I accept the claim, heard so much in central banking circles, that more accountability will harm the central bank—as long as the bank is independent. If the central bank makes good decisions, it should have no trouble explaining them to the public. If the Fed cannot articulate a coherent defense of its actions, maybe those decisions are not as good as it thinks. Indeed, being forced to articulate such a defense would probably be a good disciplinary device. Remember—and this is critical—I am talking here only about explaining the decisions after they are made, not putting them to a vote!

The Federal Reserve, tight-lipped as it is, is far from the worst offender in this regard. In fact, the Fed is probably more open and accountable than most central banks in the world. But the competition in this league is not very stiff—I think the New York Jets could win the championship in this particular league—and I believe the Federal Reserve could and should go much further. After all, we live in the most open society on the face of the earth, so just to say that we've beaten the world average is no great achievement for Americans.

The fifth ingredient in my democratic stew is that the leaders of the central bank should be politically appointed by the President, as is the current practice. When I went to the Federal Reserve Board in June 1994 as the first appointee of President Clinton, I joined Alan Greenspan, Mike Kelley, and John LaWare, who were originally sent there by President Reagan, and Larry Lindsey and Susan Phillips, who were appointed by President Bush. None of us was ever elected to anything. But Bill Clinton and George Bush and Ronald Reagan were. We obtained our political legitimacy from the men who appointed us, and they in turn got it the old-fashioned way—directly from the voters. That is as it should be.

Finally, the sixth ingredient—which I would argue should be present, but very rarely used: Central bank decisions should be reversible by the political authorities, but only under extreme circumstances. Reversal should not be routine occurrences. As I've mentioned already, a Federal Reserve decision on monetary policy can, in principle, be overturned by an act of Congress. And

Fed governors can be removed from office for good cause. These mechanisms have never been used in the history of the Federal Reserve; but America is wise to have them in place nonetheless. Delegated authority should be retrievable, not absolute.

A SUMMING UP

So, in summary, let us review how the Fed, or any other central bank for that matter, can best serve its nation with monetary policy.

To begin with, the central bank must always remember that it exists as a public institution chartered to serve the broad national interest, not the parochial interests of either the banking industry or the bond market. Often those interests coincide. But when they clash, the central bank should not hesitate before taking sides.

The highest calling of the central bank is to help stabilize the national economy. For if the bank should fail at this task, no one else will be around to pick up the pieces. In its role as macroeconomic steward, the Fed, I believe, should pursue two goals—both low unemployment and low inflation—not just one. That is what the people want and, in my view, the people have got it right.

A central bank can perform its monetary policy role better if it's independent from political manipulation, and that's probably why more and more governments around the world are granting their central banks independence these days.

Even though the Fed's independence looks superficially undemocratic, I believe it is consistent with democratic theory for several reasons: it is based on authority delegated by Congress; the basic goals of monetary policy are set legislatively; the leaders of the Fed are appointed by the President; and Congress retains ultimate control in case of dire emergency. But a central bank in a democracy has a duty to level with the public it serves, not to obfuscate. I used to ask some of my colleagues on the Federal Reserve staff in Washington what they would have thought if their father, every time he spanked them, had only said that he was doing it "to promote sustainable non-inflationary growth"—and nothing more. I don't believe that would have been considered good parenting, and I don't think it's good central banking. More fulsome explanation is appropriate.

A great Virginian, probably the greatest Virginian, once wrote, "Governments are instituted among men" (I'm sorry it was only men in those days) "deriving their just powers from the consent of the governed." It is very hard for the governed to give their consent if they don't have a clue about what is going on. Openness, accountability, and credibility are therefore, in my view, moral corollaries of central bank independence.

Furthermore, and finally, I dispute the notion that is so popular in some circles that monetary policy is best done amidst mystery, blue smoke, and mumbo jumbo. Central banks work their will through financial markets, and economists rarely argue that markets function better when they are less well informed. In my view, some small portion of the prodigious uncertainties over the effects of monetary policy exists because the markets have a hard time divining the Fed's intentions. This particular source of uncertainty can, and in my opinion should, be removed by greater openness. But that, I'm afraid, is a story for another lecture and another day.

Firewalls

John R. Walter

[R]egulations may, no doubt, be considered as in some respects a violation of natural liberty. But those exertions of the natural liberty of a few individuals, which might endanger the security of a whole society, are, and ought to be, restrained by the laws of all governments; of the most free, as well as of the most despotical. The obligation of building ***party walls, in order to prevent the communication of fire****, [emphasis added] is a violation of natural liberty exactly of the same kind with the regulations of the banking trade which are here proposed.*

— A discussion of restrictions on bank note issuance in Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*

In his 1776 *Wealth of Nations*, Adam Smith likened the regulation of bank note issues to the construction of walls that prevent the spread of fire. Smith's application of the notion of firewalls to banking seems remarkably prescient, for the subject of firewalls has surfaced repeatedly in recent discussions of proposals to reform banking. In such discussions firewalls refer to statutory and regulatory limitations on financial transactions between banks and their affiliates.[1] These limitations are analogous to fire-proof barriers in that they are meant to prevent the spread of financial difficulties within a banking company. Specifically, the restrictions should prevent a banking company from shifting financial losses from its nonbank subsidiary to its insured bank subsidiary and potentially to the federal deposit insurance fund.

This article benefited greatly from suggestions from R. Alton Gilbert (St. Louis Federal Reserve Bank), Marvin Goodfriend, Thomas Humphrey, Anatoli Kuprianov, Jeffrey Lacker, Elaine Mandaleris, James McAfee, Edward S. Prescott, Roy Webb, and John Weinberg. Alex Mendoza provided excellent research assistance. The views expressed herein are the author's and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Banking parlance employs the term firewall to refer to other types of restrictions placed between a bank and its affiliates. As discussed in this article, however, firewalls refer only to restrictions on transactions.

The current interest in firewalls has its roots in the banking crisis of the early 1930s. Passed in response to the crisis, the Banking Act of 1933 (also known as the Glass-Steagall Act) prohibited commercial banks from conducting investment banking activities and from being affiliated with investment banking firms.[2] At the time, congressmen argued that the prohibitions were necessary (1) to prevent deposits from funding stock market speculation, (2) to limit conflicts of interest, and (3) to protect bank safety.

Many bankers, regulators, and legislators now argue that these concerns are unjustified or that they can be mitigated by regulations allowing more integration of commercial and investment banking. During 1995 and 1996, Congress considered various proposals to expand the investment banking activities allowed to banking organizations by reforming the Glass-Steagall Act. Under the proposals, investment banking activities would be conducted in a subsidiary separate from the insured bank, thus isolating the bank from investment banking losses. Besides Glass-Steagall reform, each proposal included firewalls restricting transactions that might be used to shift the investment bank's losses to the affiliated bank. No proposal gathered sufficient support for enactment, but discussions continue on the issues of Glass-Steagall reform and suitable firewalls for expanded activities.

For many students of banking, the subject of firewalls raises certain pertinent questions: What are firewalls? Why do we have them? How are they enforced? In what situations are they useful? This article answers these questions. Section 1 examines firewalls. It discusses their origins and development. It examines Congress's and regulators' stated purposes in establishing such firewalls. When firewalls were initially enacted as a provision of the Glass-Steagall Act, Congress provided little explanation. Later, congressional and regulatory pronouncements clarified the goal of the restrictions: to prevent banks from undertaking transactions with nonbank affiliates on terms disadvantageous to the bank. The concern was that these transactions might be intended to rescue a troubled affiliate or to drain the bank for the benefit of its affiliate. The banking crisis of the early 1930s and later notable bank failures contributed to the concern. One such failure, discussed below, occurred in 1953 when a Dallas company, owning a chain of nonbank loan companies, purchased two small banks and proceeded to unload the loan losses of its nonbanks on these two banks. Both banks were soon closed by auditors. Ultimately, one of the banks failed, requiring FDIC payments to protect its depositors. Congress extended the firewalls in response to the incident. Section 2 describes firewall enforcement by the federal banking supervisors.

[2] Some authors use the title Glass-Steagall Act to refer only to those sections of the Banking Act of 1933 dealing with the separation of commercial and investment banking. Others use the title to refer to the entire act. I will employ the latter usage.

Incentives that could motivate a banking company to engage in such transactions receive little discussion in the legislative history of the firewalls, in regulatory commentaries on the firewalls, or in articles discussing the firewalls.[3] Without a clear understanding of a banking company's incentive, one finds it difficult to justify firewalls or to evaluate firewall recommendations included in reform proposals. Section 3 discusses several situations in which a banking company could profit from shifting a loss from its nonbank subsidiary to its bank subsidiary or from shifting bank income to a nonbank subsidiary. For example, in certain circumstances the banking company could employ such shifts to transfer wealth from the deposit insurance fund to itself. It is precisely in these situations that the firewalls are most likely to be useful.

1. THE FIREWALLS

Firewalls limit, prohibit, or set standards for transactions between banks and affiliated nonbanking companies. Sections 23A and 23B of the Federal Reserve Act prescribe firewalls and were enacted as amendments to the act in 1933 and 1987, respectively.[4] Also prescribing firewalls are the orders under which the limited-function securities subsidiaries of bank holding companies (so-called "section 20 subsidiaries") operate. In general terms, the firewalls apply to any financial transactions between a bank and its nonbank affiliates, transactions that might be used to shift the bank's resources to the nonbank.[5] For example, firewalls limit loans made by a bank to its affiliates. Such requirements are intended to limit opportunities for the bank to subsidize its affiliate with lower-priced or more risky loans than made to borrowers not affiliated with the bank.

23A and 23B Requirements

Section 23A limits banks' "covered transactions" with any single affiliate to 10 percent of the capital and surplus of the bank and with all affiliates to no more than 20 percent. Covered transactions include loans to affiliates, investments in securities issued by affiliates, purchases of assets from affiliates, acceptance of securities issued by affiliates as collateral on loans, and guarantees for affiliates (for example issuing letters of credit on behalf of affiliates). Section

[3] Articles by Gilbert (1988) and Keeton (1990) discuss firewalls peripherally. Both articles enumerate several incentives that could motivate banking companies to undertake such transactions. Macey and Miller (1992), pp. 376–77, mentions one such incentive but provides little discussion.

[4] Sections 23A and 23B are found in the United States Code at 12 U.S.C. 371c and 12 U.S.C. 371c-1, respectively.

[5] Beyond the financial transactions regulated by Sections 23A and 23B, dividend payments and stock repurchases might be used to shift a bank's resources to the parent company and on to the bank's affiliates. Dividend payments and stock repurchases are also restricted, as discussed in Wall (1984), p. 24.

23A prohibits banks from purchasing low-quality assets from affiliates. All covered transactions between a bank and an affiliate must be on terms that are consistent with safe and sound banking. Finally, all loans and guarantees must be secured by collateral equal to at least 100 percent of the value of the loans or guarantees. Transactions between banks within a bank holding company (BHC) are exempt from most firewall restrictions.[6]

Under section 23A, an affiliate of a bank is any company that controls the bank (the parent bank holding company) or is controlled by the parent company or the bank's controlling shareholders. An affiliate also can be any company in which a majority of directors are also directors of the bank or any bank subsidiary of the bank. In addition, companies that a bank sponsors and advises under contract are considered affiliates. Prime examples are the real estate investment trusts (REITs)—closed-end funds investing in real estate—whose association with banks led to bank financial losses in the 1970s (Sinkey 1979, pp. 237–55).

Section 23B expands on 23A by adding that covered transactions must be on arm's-length terms (i.e., on terms comparable to those the bank would normally offer to nonaffiliated companies). Section 23B also requires that arm's-length terms be applied to certain additional transactions that might transfer a bank's resources to its affiliate. These are the sale of securities or assets to an affiliate (section 23A covers purchases of assets from affiliates), any payments to an affiliate, any service transactions, transactions occurring when an affiliate acts as an agent or broker for the bank, or any bank transaction that indirectly benefits an affiliate.

Section 23B contains several outright prohibitions. It prohibits a bank, when acting as a fiduciary, from purchasing securities or other assets from an affiliate. For example, a bank may not purchase assets from affiliates for its trust customers. The prohibition does not apply when the fiduciary agreement specifically allows purchases of affiliate assets.[7] A bank also is prohibited from purchasing securities underwritten by an affiliate during the underwriting period. Last, section 23B prohibits a bank or its affiliate from in any way indicating that the bank is responsible for obligations of the affiliate.

[6] Prohibitions on the purchase of low-quality assets and the requirement that transactions be on terms that are consistent with safe and sound banking apply between affiliated banks (12 U.S.C. 371c(d)).

[7] Unlike most other 23A and 23B limitations intended to prevent transactions detrimental to the bank, this prohibition aims to prevent transactions detrimental to bank trust customers. The apparent concern is with the bank putting the affiliate's, and therefore the BHC's, interest ahead of the trust customer's.

Origins and Development of Sections 23A and 23B[8]

Glass-Steagall Act Firewalls

Section 23A was enacted in 1933 as part of the Glass-Steagall Act. The act was adopted in response to the stock market crash of October 1929 and the banking panic of 1932 and 1933. In 1932 and early 1933 nearly 1,850 banks failed (U.S. Congress, Senate [1933], p. 6). President Roosevelt responded by declaring a nationwide bank holiday in March 1933. In addition to limiting transactions between a bank and its affiliates, the act separated commercial and investment banking, established federal deposit insurance, made changes to the Federal Reserve System, and expanded branching privileges of national banks (Kelly 1985b, p. 41; Macey and Miller 1992, pp. 21–24).

The legislative history of the Glass-Steagall Act indicates that Congress was concerned with perceived conflicts of interest emanating from ties between commercial and investment banking, the use of bank funds for excess stock market speculation, and the adverse affects on bank safety of affiliation with securities firms (Kelly 1985a, p. 231; U.S. Congress, Senate [1933], pp. 1, 6–10). These concerns were met by prohibiting banks from directly performing investment banking functions and prohibiting banks from affiliations with investment banking companies.[9]

The legislative history of the act is fairly clear about Congress's reasons for the separation of commercial and investment banking, but it is unclear about Congress's reasons for the firewalls. The portion of the congressional committee report on the Glass-Steagall Act that specifically mentions transaction restrictions is of little help in determining why Congress believed the restrictions were necessary (see U.S. Congress, Senate [1933], pp. 9–10). The mention of the transaction restrictions occurs within a discussion of problems that can emanate from affiliations between banks and securities firms. Since bank-securities firm affiliations were, for the most part, prohibited by Glass-Steagall, there seems to have been little reason to enact transaction restrictions. Nevertheless, one

[8] See Table 1 for an outline of the history of firewall legislation and regulation.

[9] Sections 16, 20, 21, and 32 of the Glass-Steagall Act accomplish the separation of commercial and investment banking. In broad terms, and with some important exceptions, these sections may be outlined as follows:

Section 16 limits the securities-dealing activities of national banks to purchasing or selling securities on the order of their customers. In addition, national banks may not underwrite securities. Section 16 restrictions are applied to state member banks by Glass-Steagall's section 5.

Section 20 prohibits national and state member banks from being affiliated with organizations that are engaged principally in securities activities.

Section 21 prohibits firms engaged in securities activities from accepting deposits. In effect this section extends securities activity restrictions to state nonmember banks. Sections 16, 20, and 32 do not restrict these banks.

Section 32 prohibits any officer, director, or employee interlocks between national or state member banks and any organizations primarily engaged in securities activities. (Fein 1996, sect. 2.01)

Table 1 Firewalls' Legislative and Regulatory Development and Their Requirements

Banking Act of 1933 (also known as the Glass-Steagall Act)

Amended the Federal Reserve Act to add section 23A, which contained firewalls limiting transactions between Federal Reserve member banks and their affiliates. Briefly, 23A included the following restrictions:

- — Limited to a total of 10 percent of bank capital any of the following transactions: loans to an affiliate; purchases under repurchase agreement of securities held by an affiliate; investments in the obligations of an affiliate; or the acceptance of obligations of an affiliate as loan collateral. These same transactions were limited to 20 percent of the bank's capital for the sum of all affiliates.
- — All loans to affiliates were to be backed with collateral worth at least 110 percent of the value of the loan.

Bank Holding Company Act of 1956

Contained firewalls that prohibited the transactions restricted by 23A. Covered both member and nonmember banks owned by multi-bank holding companies.

Bank Holding Company Act Amendments of 1966

Repealed the BHCAct firewalls but expanded 23A to cover all insured banks, member and nonmember.

Garn-St. Germain Depository Institutions Act of 1982

Made a number of amendments to 23A, including

- — Restricted to 10 percent of bank capital all purchases of assets from any affiliate (20 percent for the sum of all affiliates); previously only purchases under repurchase agreements were restricted;
- — Added guarantees, acceptances, and letters of credit on behalf of any affiliate to the list of transactions subject to the 10 and 20 percent limitation;
- — Required that transactions be on terms consistent with safe and sound banking;
- — Prohibited the purchase of low-quality assets from affiliates;
- — Exempted affiliated banks and banks' subsidiaries from most transaction restrictions;
- — Expanded the list of assets acceptable as collateral for loans to affiliates, and modified the collateral-to-loan ratios;
- — Expanded the definition of affiliate to include REITs and other companies sponsored and advised by a bank, its subsidiaries, or affiliates.

Competitive Equality Banking Act of 1987

Amended the Federal Reserve Act to add section 23B. Section 23B requires that transactions between a bank and its affiliates must be on terms comparable to those the bank would normally offer to nonaffiliated companies (i.e., on arm's-length terms).

Financial Institutions Reform, Recovery, and Enforcement Act of 1989

Extended 23A and 23B to savings institutions.

1987 Board of Governors' Orders for Section 20 Subsidiaries

Restricted financial transactions between banks and their section 20 affiliates and between banks and customers of section 20 affiliates. Applied to section 20 subsidiaries that limit themselves to underwriting and dealing only in bank eligible securities plus commercial paper, municipal revenue securities, mortgage-backed securities, and consumer receivable-backed securities. A bank with a section 20 affiliate may not

- — make loans to issuers for the purpose of payment of principal, interest, or dividends on bank-ineligible securities underwritten by its section 20 affiliate;
- — make loans secured by or for the purpose of purchasing bank-ineligible securities underwritten by its section 20 affiliate during the underwriting period, and for 30 days thereafter;
- — make loans or provide guarantees (for example a letter of credit) that will enhance the credit-worthiness or marketability of a bank-ineligible security underwritten by its section 20 affiliate;
- — purchase bank-ineligible securities underwritten by its section 20 affiliate during the underwriting period and for 60 days after;
- — purchase from its section 20 affiliate bank-ineligible securities in which the section 20 makes a market.

1989 Board of Governors' Orders for Section 20 Subsidiaries

Restricted financial transactions between banks and their section 20 affiliates and between banks and customers of section 20 affiliates. Applied to section 20 subsidiaries that underwrite and deal in debt and equity securities. The prohibitions under 1987 orders also applied to these subsidiaries. In addition, under the 1989 orders a bank may not

- — make loans to its section 20 affiliate;
- — make any financial asset purchases from, or sell financial assets to, affiliated section 20s;
- — provide a guarantee, such as a letter of credit, for its section 20 affiliate.

can imagine at least three important factors that might have led Congress to enact the transaction restrictions.

First was the failure of the Bank of the United States in December 1930. This failure was caused by the bank president's appropriation of large portions of bank funds, through the bank's affiliates, to his own personal and highly speculative business ventures. His actions created widespread suspicion of all commercial bank affiliates (Perkins 1971, pp. 496–97).

Second, 1931 subcommittee hearings leading to the Glass-Steagall Act included a discussion of a number of securities affiliate "abuses" Congress may have believed applicable to other types of affiliates (U.S. Congress, Senate [1931], pp. 1052–68). Specifically, the hearing report notes that a bank might extend credit to a troubled affiliate to rescue it, purchase assets of a troubled affiliate, or make unsafe loans to the affiliate's customers. According to the report, "[w]hen dealing with its affiliate, the bank is really dealing with itself." As a result, "there tends to be a breaking down of those limitations on the extension of credit which the bank sets up in other cases to guard against the making of excessive or poorly secured loans" (U.S. Congress, Senate [1931], p. 1066).

Third, in spite of the general prohibition of bank-securities firm affiliation, the Glass-Steagall Act left banks with the ability to affiliate with firms with limited securities powers. Perhaps Congress foresaw similar abuses with these affiliates and wanted to restrict transactions between banks and these limited securities affiliates.

For whatever reason or reasons the act introduced firewalls for most types of bank affiliates. According to a Senate Report on the bill that became the Glass-Steagall Act, an important goal of the bill was "[t]o separate as far as possible national and member banks from **affiliates of all kinds**" [emphasis added] (U.S. Congress, Senate [1933], p. 10). Section 23A firewalls helped accomplish the goal.

As passed in 1933, section 23A applied only to Federal Reserve member banks. Included in this category are all national banks, meaning banks chartered by the federal government and thus required by statute to be Federal Reserve members, and state-chartered banks that elect to become members of the Federal Reserve.

Firewalls Extended by the Bank Holding Company Act of 1956

The Bank Holding Company Act of 1956 (BHCAct) extended firewalls similar to those in 23A to all banks, member and nonmember, owned by multi-bank BHCs.[10] The firewalls extension was at least in part provoked by the failure of a Chicago bank, which stemmed from transactions with its parent BHC (discussed below). The BHCAct included a provision completely prohibiting

[10] The BHCAct of 1956 covered only BHCs owning two or more banks. In 1970 the act was amended to include one-bank BHCs.

specified transactions: bank loans to, investments in, and purchases of assets from affiliates or the parent BHC (Senate Rep. No. 1095, 84th Cong., Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 2482, 2496–97 [1956]). Section 23A firewalls differed in that they allowed such transactions as long as their dollar aggregate amounted to no more than 10 or 20 percent of bank capital.

For the most part, the BHCAct (of which the firewall provisions were only a small portion) was the result of congressional concern over the lack of regulatory control of bank holding companies (Senate Rep. No. 1095, 84th Cong., Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 2482, 2483 [1956]). While banks were extensively regulated and supervised either under federal or state law, bank holding companies were relatively free of such regulation prior to the passage of the act. The act brought the activities of multi-bank BHCs under Federal Reserve regulatory and supervisory authority.

Just as with the reports on section 23A, the House and Senate committee reports on the BHCAct lack explanation. In this case, they fail to explain Congress's motive for completely prohibiting specified transactions rather than simply limiting them, as did section 23A. Yet, in contrast to the legislative history surrounding the passage of 23A firewalls, the reports do fairly clearly explain Congress's motivation for the firewalls in general: to prevent a parent BHC from taking "undue advantage of the resources of its subsidiary banks" (Senate Rep. No. 1095, 84th Cong., Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 2482, 2496 [1955]). While the history indicates that "no widespread abuse of this nature has been brought to the attention of [Congress]," one such case was discussed in hearings leading to the passage of the act. In this case, a Dallas company that owned a chain of (nonbank) small-loan companies purchased slightly more than 50 percent of the shares of two small Chicago banks. Following the bank purchases, the Dallas company sold to the banks questionable assets held by the BHC at face value less a small discount (2 percent). The asset sales shifted BHC losses to the banks, and within a few weeks of the shift the banks were closed by state auditors. Ultimately, some of the shifted losses were borne by the FDIC, as the FDIC was able to reopen one of the banks only after providing financial assistance (U.S. Congress, House [1955], pp. 18–19; FDIC 1953, p. 8). According to a House report discussing the BHCAct, failure to institute the BHCAct restrictions on interaffiliate transactions "would be to invite a repetition of the [Dallas BHC incident]" (U.S. Congress, House [1955], p. 19).

Firewalls Further Extended by 1966 BHCAct Amendments

In 1966 the Bank Holding Company Act was amended to repeal the act's prohibitions of interaffiliate transactions, once again leaving only the firewalls laid out in section 23A to restrict such transactions. At the same time, 23A was amended specifically to cover transactions between a bank and its parent

BHC and subsidiaries of its parent. In addition, the Federal Deposit Insurance Act was amended to apply section 23A to all federally insured commercial banks, subjecting almost all banks to the requirements of section 23A (Rose and Talley 1982, p. 693; Senate Rep. No. 1179, 89th Cong., Sess. 2, reprinted in *U.S. Code Congressional and Administrative News*, 2385, 2396 [1966]).

Garn-St. Germain Act of 1982 Amendments to 23A

In 1982 section 23A was amended in several important ways. According to a Senate Banking Committee report, the amendments had three objectives. First, the changes were meant to liberalize unnecessarily restrictive provisions in the statute. Second, they were to close several loopholes. Third, they were intended to reorganize and clarify the statute to improve compliance and enforcement (Senate Rep. No. 97–536, 97th Cong., Sess. 2, reprinted in *U.S. Code Congressional and Administrative News,* 3054, 3085 [1982]).

Several significant circumstances preceded, and ultimately led to, the 1982 revisions to section 23A. For a number of years bankers had argued that section 23A unnecessarily stifled low-risk, interaffiliate transactions. In addition, in the mid-1970s several large banks were adversely affected by transactions with affiliates or with firms the banks sponsored and advised.

The failure of Hamilton National Bank in 1976 served to illustrate how section 23A might fail to prevent interaffiliate transactions detrimental to the bank.[11] The failure of the Chattanooga-based, $461-million-asset Hamilton National Bank was the third largest bank failure in U.S. history up to that time. Hamilton National Bank was owned by a BHC, Hamilton Bancshares, Inc., which also owned nonbank subsidiaries including Hamilton Mortgage Corporation, an Atlanta-based mortgage banking company specializing in real estate development loans. In 1974, the real estate industry experienced major problems. Consequently, growing numbers of Hamilton Mortgage Corporation's borrowers began to default. When examined in September 1974, Hamilton National Bank was holding more than $100 million in mortgages originated by the mortgage subsidiary, many of which were troubled (Comptroller of the Currency 1977, p. 233). Ultimately the bank failed largely from losses on loans purchased from its mortgage affiliate (Comptroller of the Currency 1977, pp. 233–34; Sinkey 1979, pp. 199–205).[12]

[11] Hamilton National Bank was the only one of 120 bank failures occurring in the ten years prior to the 1982 amendment to section 23A that could be directly attributed to interaffiliate asset transactions (Morgan Guaranty [1983] as cited in Saunders [1988], p. 168, footnote 43).

[12] Hamilton National Bank's purchases from its affiliate far exceeded 10 percent of its capital. While at the time 23A did not specifically restrict a bank's purchase of loans from its affiliate, the statute defined "extensions of credit" to an affiliate, which were restricted, to include a bank's "discount of promissory notes" held by its affiliate. Under a 1974 Board of Governors interpretation of 23A, the phrase "discount of promissory notes" includes the purchase of loans. This interpretation indicates that, with the exception of the amount permitted under the 10 percent limitation, Hamilton's purchases were in violation of 23A (Board of Governors 1974, pp. 726–27).

Also during the mid-1970s, several large U.S. banks sustained significant losses as a result of bailing out associated REITs. A number of large banks sponsored and advised these closed-end real estate investment funds in the 1970s. As the real estate market slumped in 1974 and REITs experienced financial problems, REITs' ability to issue commercial paper diminished. Because commercial paper issues provided a major portion of REIT funding, the REITs faced insolvency. To prevent REIT failures, a number of banks gathered together to fund a costly rescue operation (Sinkey 1979, pp. 237–55; Cornyn et al. 1986, pp. 187–88).

In response to concern with perceived deficiencies in section 23A, Congress called on the Federal Reserve to recommend amendments (Rose and Talley 1983, p. 424). The Fed's proposals served as the basis for the Banking Affiliates Act of 1982, which made a number of substantive changes to section 23A (Rose and Talley 1983, p. 424; Senate Rep. No. 97–536, 97th Cong., Sess. 2, reprinted in *U.S. Code Congressional and Administrative News,* 3054, 3085 [1982]). The Banking Affiliates Act was enacted as a section of the Garn-St. Germain Depository Institutions Act of 1982. Guiding the proposals was the view that the statute was meant to prevent the misuse of a bank's resources stemming from large-scale transactions with affiliates (Board of Governors 1981, pp. 1, 6, 28, 31, 39, 42–43). Specifically, the Fed's concern with interaffiliate transactions was that "because of this relationship [affiliation], the bank may engage in transactions that may adversely affect its condition. Such transactions may be designed either to rescue a financially troubled affiliate or to 'drain' a bank for the benefit of an affiliate" (Board of Governors 1981, p. 6).

The Federal Reserve proposed that REITs and other companies sponsored and advised by a bank or a bank's subsidiary or affiliate be considered affiliates. This proposal was motivated by REIT losses and the Fed's view that the relationship between a bank and a REIT it sponsors is in some respects an affiliate relationship (Board of Governors 1981, pp. 23–25). Other important changes to section 23A were a broadened definition of asset purchases considered "covered" transactions; the addition of guarantees, acceptances, and letters of credit issued on behalf of an affiliate as "covered" transactions; the addition of a provision requiring that all "covered" transactions with affiliates be on terms that are consistent with safe and sound banking practices; the prohibition of purchases of low-quality assets; the exemption of affiliated banks from most restrictions on transactions; the exclusion of banks' majority-owned subsidiaries from the definition of an affiliate; and an expanded list of assets that are acceptable as collateral backing transactions with affiliates (Senate Rep. No. 97–536, 97th Cong., Sess. 2, reprinted in *U.S. Code Congressional and Administrative News*, 3054, 3085–86 [1982]).

CEBA and FIRREA Augment 23A Firewalls

The Federal Reserve Act firewalls were expanded in 1987 with the addition of section 23B, passed as a provision of the Competitive Equality Banking Act of 1987 (CEBA). Previously in 1983 and 1984, Senate bills had proposed the expansion of bank securities, insurance, and real estate development powers and included proposals to amend the Federal Reserve Act by adding section 23B. Although these bills were not adopted, 23B resurfaced a few years later in CEBA (Miles 1988, p. 478, footnote 4). The Financial Institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA) extended sections 23A and 23B to savings institutions following the thrift industry crisis of the 1980s (U.S. Congress, House [1989], pp. 169–70; 12 U.S.C. sec. 1468(a)).

Firewalls Surrounding Underwriting Subsidiaries

A number of large BHCs currently take advantage of a loophole in the Glass-Steagall Act and operate securities-underwriting subsidiaries (named "section 20 subsidiaries" for the section of the act containing the loophole).[13] As discussed below, the loophole is limited, so securities subsidiaries of BHCs are somewhat circumscribed compared to securities firms not owned by BHCs. BHCs' underwriting subsidiaries face thicker interaffiliate transaction firewalls, instituted by Board of Governors orders, than those found in sections 23A and 23B. The thicker firewalls have the same purpose as the 23A and 23B firewalls.

Before the Glass-Steagall Act was passed in 1933, member banks could be affiliated with securities-underwriting companies. Following its passage, and continuing until the 1980s, member banks were denied such affiliations. Nevertheless, under the Glass-Steagall Act banks and their affiliates were allowed to underwrite certain assets such as U.S. government securities and general obligation municipal securities (Fein 1996, pp. 8-7, 8-12). Such securities are known as "bank-eligible" securities. With the advent of significant banking deregulation and heightened competition in the 1970s and 1980s, banking companies began to argue for the authority to conduct additional securities activities. In late 1984, Citicorp submitted an application to the Federal Reserve Board for approval to begin underwriting certain ineligible securities in its eligible securities subsidiary. While this initial application was withdrawn, Citicorp resubmitted a modified application in 1985 and was later joined by two other large New York BHCs making similar applications.

In response to these applications, the Federal Reserve Board reviewed the Glass-Steagall Act and concluded that the law does not prohibit a company affiliated with a member bank from underwriting and dealing in certain

[13] As of 1996 there were 38 section 20 subsidiaries of BHCs (*American Banker,* May 10, 1996, p. 3).

securities if these activities provide no more than 5 percent of the gross revenues of the subsidiary (Board of Governors 1987, p. 476; Fein 1996, p. 8-16). Glass-Steagall does contain sections which prevent banks from conducting investment banking activities directly and prevent member banks from affiliating with companies that do so. Section 20 prohibits Federal Reserve member banks from affiliation with any firm "engaged principally" in the issue, flotation, underwriting, public sale, or distribution of securities. The Board ruled that if revenues from these prohibited activities account for no more than 5 percent of the subsidiary's total revenue, the subsidiary is not "engaged principally" in the activity. In this case the subsidiary may be affiliated with a member bank without violating the Glass-Steagall Act (Board of Governors 1987, pp. 475–76).[14] The only other securities beyond eligible securities that these subsidiaries were authorized to underwrite and deal in were commercial paper, municipal revenues securities, mortgage-backed securities, and consumer receivable-backed securities (Fein 1996, p. 8-49).

1987-Order Firewalls

In 1987, the Board of Governors of the Federal Reserve established additional firewalls that not only prohibited certain transactions between a bank and its affiliated section 20 subsidiary, but also prohibited them between a bank and the customers of its section 20 affiliate. The Board's 1987 firewalls are provisions of Board orders authorizing securities underwriting and dealing. A bank with a section 20 affiliate may not

— make loans to securities issuers for the purpose of payment of principal, interest, or dividends on bank-ineligible securities underwritten by its section 20 affiliate;

— make loans secured by or for the purpose of purchasing bank-ineligible securities underwritten by its section 20 affiliate during the underwriting period, and for 30 days thereafter;

— make loans or provide guarantees (for example a letter of credit) that will enhance the creditworthiness or marketability of bank-ineligible securities underwritten by its section 20 affiliate;

[14] Section 20 of the Glass-Steagall Act applies only to banks that are members of the Federal Reserve System. As previously noted, this contingent includes all national banks, i.e., those chartered by the federal government, and state-chartered banks that choose to be members (state members). Since section 20 does not apply to state-chartered nonmember banks, these banks may be affiliated with securities-underwriting companies engaged principally in underwriting ineligible securities. A nonmember bank may not engage in underwriting in the bank itself because Glass-Steagall's section 21 prohibits such activities for member and nonmember banks. According to a 1981 Supreme Court ruling, section 21 does not extend to bank subsidiaries or affiliates (Fein 1996, p. 2-6). Nonmember banks are subject to the supervision of the FDIC, which has established several firewalls similar to those established by the Federal Reserve for section 20 affiliates (Fein 1996, pp. 8-64.3–8-68; U.S. General Accounting Office 1995, pp. 74–75).

— purchase bank-ineligible securities underwritten by its section 20 affiliate during the underwriting period and for 60 days after;
— purchase from its section 20 affiliate bank-ineligible securities in which the section 20 makes a market.[15]

In its discussion of the firewalls, the Federal Reserve indicated that the prohibitions on loans are meant to prevent "imprudent" and "unwise" lending to help out affiliated section 20s (Board of Governors 1987, pp. 496–97). Likewise, the securities purchase prohibitions are meant to preclude "unwarranted" and "detrimental" purchases by banks intended to prevent section 20 losses (Board of Governors 1987, pp. 497–98).

1989-Order Firewalls

In 1989 the Fed increased the revenue limit from 5 to 10 percent and expanded the types of securities the subsidiaries could underwrite and deal in to include most debt and equity securities (Fein 1996, pp. 1-16–1-18, 8-24.2, 8-29).[16] For those section 20 affiliates that chose to expand underwriting and dealing activities beyond the four types of securities authorized in 1987, firewalls were augmented. The major change was that of prohibiting loans by banks to their section 20 affiliates. Also, under the 1989 orders, banks were prohibited from making any financial asset purchases from, or selling such assets to, affiliated section 20s.[17] Last, banks were prohibited from providing a guarantee, such as a letter of credit, for their section 20 affiliate. The stricter firewalls for section 20s were developed out of concern for the increased potential for losses from the expanded operations allowed to subsidiaries operating under the 1989 orders (Board of Governors 1989, pp. 203–06). Because of these firewalls and other more demanding restrictions, some section 20 affiliates have chosen to confine

[15] These prohibitions contain certain exceptions. For a complete list of all underwriting subsidiary firewalls, see Board of Governors, *Bank Holding Company Supervision Manual* (1995), sect. 2185.0, pp. 15–28, or U.S. General Accounting Office (1995), pp. 70–73. Additional requirements and restrictions are placed on relations between banks and underwriting affiliates beyond these transactions firewalls. The additional requirements are meant to limit the opportunity for the exploitation of conflicts of interest and to provide a clear separation between the underwriting and bank subsidiaries of the BHC. For example, BHCs are required to maintain separate offices for the section 20 subsidiary and limit employee, officer, and director interlocks and communication of confidential customer information between section 20 subsidiaries and affiliated banks. A section 20 subsidiary is also required to disclose that it is separate from any affiliated bank and that securities offered, sold, or recommended are not FDIC insured (Board of Governors, *Bank Holding Company Supervision Manual* 1995, sect. 2185.0, pp. 15–28).

[16] In July 1996 the Board of Governors proposed an increase in the revenue limit to 25 percent (Board of Governors 1996b).

[17] The 1989 orders granted an exception allowing banks to purchase U.S. Treasury securities from, and sell such assets to, section 20 affiliates. In October 1996 the exception was expanded to include all securities with readily identifiable and publicly available market quotations (Board of Governors 1996a).

their activities to those allowed by the 1987 orders so that they are limited only by the 1987 firewalls.

2. ENFORCEMENT

Sections 23A and 23B and the section 20 firewalls are enforced by bank and bank holding company examiners in periodic on-site examinations. In broad terms, examination for firewall compliance follows a simple two-step pattern. Examiners first determine if the bank or affiliate has procedures in place that should prevent violations. Second, they run tests to determine if those procedures are being followed.

Banks are examined for compliance with 23A and 23B firewalls during each periodic safety and soundness examination. Bank examiners first obtain documents describing the bank's policies and procedures to verify that the bank has rules in place to avoid exceeding section 23A limits and to ensure that it is meeting the requirements of sections 23A and 23B. To test whether the bank's rules are observed, the examiner obtains from the bank a list of all transactions between the bank and affiliates. This list is verified against other sources to ensure that the bank did not exclude transactions. For example, to test whether he has a complete list of interaffiliate loans, the examiner will compare the list supplied by the bank with other accounting records maintained by the bank, such as customer liability records. Using the verified list, the examiner reviews the transactions to ensure (1) that together they do not exceed limits established by section 23A, (2) that 23A collateral requirements are met, (3) that low-quality loans have not been purchased from affiliates, and (4) that all transactions with affiliates are consistent with safe and sound banking practices (Board of Governors, *Commercial Bank Examination Manual* 1995).

Likewise, all bank holding companies are examined, or inspected, periodically. Only examiners from the Federal Reserve conduct these inspections. The holding company's policies and procedures regarding transactions between subsidiary banks and affiliates are reviewed to verify that procedures are in place to prevent violations of 23A and 23B (Board of Governors, *Bank Holding Company Supervision Manual* 1995, sect. 2020.1, p. 6, Item 3(g)). BHC inspections include a "thorough analysis of most intercompany transactions" to test whether the BHC's stated policies are being followed (Board of Governors, *Bank Holding Company Supervision Manual* 1995, sect. 2020.0, p. 2). The analysis includes a review of a list of all transactions between subsidiary banks and affiliates and of the terms, conditions, and circumstances of each listed transaction to ensure that none violates 23A or 23B.

At least once a year, section 20 subsidiaries are examined by Federal Reserve examiners. As with bank and BHC examinations, the section 20 examiner reviews procedure manuals to ensure they provide guidelines that will prevent firewall breaches. For example, the examiner may obtain copies of affiliated banks' loan policy manuals to verify that they clearly state the types of loans restricted, such as loans to the section 20's current underwriting customers. The examiner will also assure that there is a procedure in place for notifying loan officers in the affiliated bank of the current underwriting customers of the section 20 subsidiary. The examiner may then perform tests to verify consistency with the procedures. For example, to determine that the bank has not financed the purchase of securities underwritten by the affiliated section 20 unit, the examiner will obtain a list of securities underwritten by the section 20 entity and compare the list to the bank's collateral ledger. The ledger lists the collateral against which loans are made, allowing the examiner to confirm that there is no overlap (Board of Governors, *Bank Holding Company Supervision Manual* 1995, section 2185.0, p. 13).

The responses of supervisory agencies to violations vary depending upon the severity of the violation. Examiners may simply notify the bank or the BHC of the violation and require the modification of procedures to prevent its recurrence. The supervisory agencies can impose penalties including cease and desist orders; actions to remove directors, officers, and employees; and monetary penalties on institutions or individuals (Spong 1995, pp. 113–17; Board of Governors, *Bank Holding Company Supervision Manual* 1995, sect. 2110.0). For example, in 1983 the FDIC assessed a $1,000 penalty against a bank director, in part because of violations of 23A. The bank's board of directors, with this director's concurring vote, had authorized the bank to make loans to an affiliate above 23A loan limits, backed with collateral insufficient to meet 23A requirements (*Fitzpatrick v. Federal Deposit Insurance Corp*., 765 F.2d 569 [6th Cir. 1985]).

3. WHEN FIREWALLS ARE USEFUL

The preceding historical review reveals that firewalls were prescribed to prevent adverse resource shifts. What has never been adequately explained, however, is why a BHC would want to effect such shifts in the first place. For a BHC could only benefit from the shifts if they reduced losses or enhanced profits.[18] In most

[18] Some commentators argue that banks may have a lower cost of funds than nonbanks due to federal deposit insurance. Allowing banks to lend to their nonbank affiliates, thus passing on the deposit insurance subsidy to the nonbank, could give companies that own both a bank and a nonbank an unfair competitive advantage over companies not owning banks (for example see Macey and Miller [1992], p. 377, or the discussion of protestants' argument against greater securities powers for BHCs in Board of Governors [1987]). Firewalls might be useful for limiting or preventing the grant of this unfair advantage and associated allocative efficiency losses. But there is reason to question the quantitative significance of the deposit insurance subsidy, and therefore banks' ability to subsidize their nonbank affiliates. If the subsidy were significant, one would expect banks to displace nonbank sources of credit over time. Instead, over the last several decades banks have lost share in the funds market, or at best held stable (Boyd and Gertler [1994], p. 2). Further, a BHC can benefit from shifting the subsidy from its bank to its nonbank affiliate only if profit opportunities in the nonbank affiliate, without the benefit of subsidized funds, exceed profits in the bank.

cases, simply shifting losses from one subsidiary to another generates no such benefit, and the firewalls would be unnecessary. Still, there are two cases where the BHC might profit from shifts such that firewalls would be necessary. In case number one, some commentators have argued that BHCs can be expected to use bank resources to prevent the failure of a troubled nonbank subsidiary, thus preserving the reputation of the BHC.[19] That is to say, shifts can reduce BHC losses when the BHC's reputation would be damaged by the failure of a nonbank subsidiary. In case number two, a BHC can reduce losses by shifting them from a highly capitalized to a less capitalized subsidiary. The reason? Such shifts allow the BHC to escape some of the loss by taking advantage of shareholders' limited liability. In this case, if the less capitalized subsidiary is an insured bank, then the reduction of loss amounts to a transfer of wealth from the deposit insurance fund to the BHC. Firewalls, and their associated penalties, may be useful in reducing the likelihood of shifts motivated by these incentives since, in either case, nonbank losses can be shifted to the deposit insurance fund.

To illustrate why, in most cases, there is no benefit when losses are shifted, consider the example of a bank and a commercial finance company owned by a BHC. Should the finance subsidiary expect to incur loan losses amounting to $10 million, the BHC would be made no better off by shifting the potential loss to the bank.[20] The BHC, whose assets include the net worth of both the bank and nonbank, can expect to suffer an equivalent $10 million decline in the value of its total assets whether the bank or the nonbank subsidiary bears the loss. Firewalls aside, the bank or the BHC, because of minimum capital requirements, is likely to face strictures imposed by examiners in response to any shifts of losses or income that lower the bank's capital. Accordingly, the BHC not only gains nothing, it may suffer some cost from any shift of losses. In this case, a shift of losses is not profitable or attractive, with or without firewalls.

On the other hand, as mentioned earlier, the story could be different when a BHC's reputation is at stake. Some authors argue that a BHC has an incentive to use all its resources, including resources of its bank subsidiaries, to support troubled nonbank subsidiaries (see, for example, Cornyn et al. [1986], or Keeley and Bennett [1988]). The major factor cited is the desire to preserve the reputation of the BHC. Cornyn et al. argue that "experience suggests that BHC management will draw on the financial resources of the bank to assist

---

[19] This section assumes that BHC managers will seek the best interest of shareholders; in other words, that they are profit maximizers. Gilbert (1988, pp. 70–71) argues that the desire of BHC managers to save their jobs, when confronted with a troubled affiliate and managers' fraudulent schemes to steal from banks, can motivate such shifts.

[20] This assumes that the bank's capital is at least $10 million. As will be discussed later, the BHC can benefit by shifting the loss if it exceeds bank capital.

a troubled subsidiary" (p. 191).[21] The primary reason is the "avoidance of serious reputation damage to the holding company and its bank" (p. 187) from the failure of a subsidiary. They cite several cases, one of which saw banks taking on some of the losses of REITs, discussed earlier in this article.[22]

If a BHC owns a nonbank with expected losses exceeding the nonbank's capital, then the BHC may be best served by simply allowing the subsidiary to fail. Because of the protection afforded the BHC by shareholders' limited liability, the BHC stands to lose no more than the value of its equity investment in the subsidiary. The subsidiary's creditors suffer the remainder of the loss. But the failure of one subsidiary may affect the public's perception of the competence of the BHC management and of the soundness of other segments of the company. In turn, the shift in public perception could lead to creditor demands for higher interest rates from the BHC and its surviving nonbank subsidiaries. In addition, any harm to the reputation of surviving subsidiaries may mean lost business opportunities. Bailing out the troubled nonbank subsidiary with bank resources, rather than allowing it to fail and seeking refuge in limited liability, might leave the BHC's shareholders better off, even though it leaves the bank worse off. If the bank is made significantly worse off, the probability of a deposit insurance claim is increased.

Consider the example of a BHC owning three nonbank subsidiaries, each having net worth (owners' equity) of $10 million (see Table 2, line 1). Additionally, it owns a bank with net worth of $100 million. If one of the subsidiaries faces a loss of $16 million, the "failure" cost to the BHC is $10 million—the amount of its lost owners' equity in the troubled subsidiary should it be allowed to fail. In addition to the failure cost, the BHC will face "reputation costs" (T-accounts in Table 2 list reputation as an intangible asset of each subsidiary and reputation cost as a decrement to the asset) for the sake of this example, summing to $8 million (Table 2, line 2). Such costs include increased demands from the surviving subsidiaries' creditors and lost business opportunities.[23] Consequently, the total cost to the BHC of allowing the failure of the nonbank subsidiary amounts to $18 million—failure costs of $10 million plus reputation costs of $8 million. If instead $7 million of the nonbank's loss were shifted to the bank with the bank purchasing some of the nonbank's troubled assets at greater than their market value, for example—total BHC losses would sum to $16 million (nonbank's losses of $9 million plus bank losses of $7 million). The BHC's shareholders would be better off

[21] Gilbert (1988), pp. 70–71, also discusses preservation of reputation as a motive for transferring bank resources to the nonbank.

[22] Cornyn et al. (1986) conclude that firewalls alone are not sufficient to prevent BHCs from drawing on bank resources to assist troubled affiliates.

[23] In my simplified example I assume the parent BHC issues only equity, so only the subsidiaries have creditors.

Table 2

$ Millions

(1)

Parent BHC*

Assets	NW
Assets 130	NW** 130

Nonbank A		Nonbank B		Nonbank C		Bank	
TA** 96	Liab** 90	TA 96	Liab 90	TA 96	Liab 90	TA 980	Liab 900
Rep** 4	NW 10	Rep 4	NW 10	Rep 4	NW 10	Rep 20	NW 100

(2)

Nonbank A bears complete loss

Parent BHC

Assets	NW
130-(10+8)	130-(10+8)

Parent's assets and net worth decline by $18M due to declines in net worth of $10M at Nonbank A, $2M each at Nonbanks B and C, and $4M at Bank

Nonbank A		Nonbank B		Nonbank C		Bank	
TA 96-16	Liab 90-10	TA 96	Liab 90	TA 96	Liab 90	TA 980	Liab 900
Rep 4-4	NW 10-10	Rep 4-2	NW 10-2	Rep 4-2	NW 10-2	Rep 20-4	NW 100-4
Nonbank A suffers $16M loss—declares bankruptcy. Loss exceeds net worth, leaving creditors with a $10M loss. Reputational assets are lost when the subsidiary fails.		Reputation of Nonbank B damaged $2M causing net worth to decline by $2M		Reputation of Nonbank C damaged $2M causing net worth to decline by $2M		Reputation is damaged $4M causing net worth to decline by $4M	

(3)

Nonbank A bears only part of loss:

Parent BHC

Assets	NW
130-(9+7)	130-(9+7)

Parent's assets and net worth decline by $16M due to $9M decline in net worth of Nonbank A and $7M decline in net worth of Bank

Nonbank A		Nonbank B		Nonbank C		Bank	
TA 96-9	Liab 90	TA 96	Liab 90	TA 96	Liab 90	TA 980-7	Liab 900
Rep 4	NW 10-9	Rep 4	NW 10	Rep 4	NW 10	Rep 20	NW 100-7
Nonbank A suffers a $16M loss, but the value of its assets falls by only $9M because $7M of its loss is shifted to Bank. Net worth falls by $9M						Value of Bank's assets falls by $7M due to shift of loss from Nonbank A. Net worth likewise declines $7M	

* Here I focus on the parent corporation. One could focus instead on the consolidated enterprise—all assets, liabilities, and net worth on one balance sheet. The results are equivalent.

** NW is net worth, TA is tangible assets, Rep is reputation—an intangible asset to the subsidiary, Liab is liabilities.

by $2 million, creating an incentive for the BHC to shift the losses (Table 2, line 3).

The reputation-defense argument for shifting losses relies upon the assumption that outsiders have difficulty ascertaining the true value of a BHC's bank and nonbank subsidiaries' assets. Given this assumption, any new information on the quality of subsidiaries' management would cause outsiders to revise their estimates of these subsidiaries' asset values. For example, the failure of one BHC subsidiary, itself new information about assets of the failed subsidiary, is likely to be taken as new negative information about the quality of the management of surviving subsidiaries of the same company. As a result, outsiders will lower their estimates of survivors' asset values. Reputation damage costs will result when outsiders lower their view of the strength of surviving subsidiaries. As noted earlier, these costs include creditors' demands for higher interest rates from survivors and survivors' loss of future business opportunities.

Accordingly, if a subsidiary's failure can be prevented both at a cost less than the BHC's estimate of reputation costs and in a manner that avoids the revelation of losses to outsiders, BHC shareholders will benefit. Precisely because bank assets are difficult for outsiders to value, losses shifted to the bank may go unnoticed.

While the parent company could assist the nonbank itself, it may be better able to hide the loss by shifting it to its bank subsidiary. If the parent holds few liquid assets, assisting the troubled affiliate could require the parent to issue debt, commercial paper, or sell one of its other subsidiaries. Such actions necessitate public disclosures and notices, highlighting the BHC's predicament. If instead the bank purchased the nonbank's troubled assets at inflated prices, this interaffiliate transaction would quite likely escape outside notice.

Nevertheless, the BHC has a counterincentive weighing against the previously discussed incentive to shift losses to its bank. Any shift of losses will increase the probability of the bank's failure. Surely, the bank's failure would have reputation consequences for the BHC, so that any increase in the probability of the bank's failure due to a shift will tend to offset the BHC's desire to shift losses. While this counterincentive may be too weak to prevent all shifts of losses, it will tend to prevent those most dangerous to bank safety.

While preservation of reputation provides one set of incentives under which a BHC might benefit from shifting nonbank losses to a bank, a second set of incentives arise because of shareholders' limited liability. When the nonbank faces a potential loss smaller than its net worth but larger than the bank's, the BHC, taking advantage of limited liability, reduces its loss by shifting it to the bank. Consider the example of a BHC owning a bank with net worth of $100 million and a real estate lending company with net worth of $200 million (Table 3, line 1). Should the real estate subsidiary expect to incur loan losses of $125 million, the BHC's management could potentially save its shareholders $25 million by shifting the loss to the bank. Such a shift could be accomplished by bank purchases of bad loans from the nonbank or bank loans made to the nonbank's troubled borrowers. As owner of the stock of both the bank and the nonbank, the BHC is liable for losses only up to the value of its ownership interest in each company, according to the principle of limited liability.[24,25] If the loss is borne by the nonbank (see Table 3, line 2), the value of the BHC's nonbank stock holdings will shrink by $125 million such that shareholders will lose that amount. If, on the other hand, the potential loss can be shifted to the bank, the shareholders' loss is limited to the BHC's ownership interest in the bank, or $100 million.[26] If the borrowers recover, the BHC neither gains nor loses from the shift. If instead, the expected $125 million loss is realized, the bank fails but the shift saves the BHC $25 million (Table 3, line 3). In the latter case, the BHC's shareholders benefit at the expense of the deposit insurance fund.

Nonbank subsidiaries are typically considerably smaller than their bank affiliates. Thus, the circumstances under which the above-mentioned tactic would benefit the BHC's shareholders might appear to be quite limited. But this situation may change if the current limitations on affiliations between banks and investment banks are relaxed. Banks might be affiliated with investment banks as large as themselves. In addition, a bank that is itself troubled may find

[24] The protection of holding companies provided by the principle of limited liability is not iron-clad. The major grounds upon which the courts will disregard limited liability is when it is determined that the corporation has engaged in conduct or made representations likely to deceive creditors into thinking the shareholder was the real debtor (Posner 1977, p. 297). For further discussion of when the "corporate veil is pierced" and holding companies are held liable for subsidiary losses beyond their equity investment, see Black, Miller, and Posner (1978), pp. 395–96, Posner (1977), pp. 296–97, and U.S. General Accounting Office (1987), pp. 19–20.

[25] Two significant examples of banking policies that depart from the principle of limited liability are the Fed's "source-of-strength" policy, and the cross-guarantee provision of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA). The Federal Reserve has attempted to require BHCs to provide financial assistance to troubled subsidiary banks (Gilbert 1991, p. 4, especially footnote 3; Keeton 1990, pp. 60–61) in its enforcement of its source-of-strength policy. In other words, it has attempted to hold a bank's shareholders (the BHC) responsible for the losses of the bank beyond their equity investment. The Fed's source-of-strength doctrine, a part of Federal Reserve Regulation Y, requires a BHC to act as a source of strength to its bank subsidiaries. Nevertheless, the ability of the Federal Reserve to enforce its source-of-strength regulation may be in some doubt. In 1988 the Fed commenced proceedings that might have required MCorp, a Texas BHC, to aid its insolvent bank subsidiaries. The attempt was repudiated in a federal appeals court. The appeals court held that the source-of-strength policy was in excess of the Fed's regulatory jurisdiction. The U.S. Supreme Court held that the lower courts lacked jurisdiction because MCorp's challenge to the Fed's proceeding had been premature, leaving the underlying substantive issue unresolved (Macey and Miller 1992, p. 656; *Board of Governors v. MCorp Financial, Inc.*, 502 U.S. 32 [1991]). FIRREA authorizes the FDIC to recover from a solvent bank losses incurred by the FDIC in connection with the failure of an affiliated bank (Keeton 1990, p. 58). Two federal circuit courts of appeal upheld FIRREA's cross-guarantee against constitutional challenges (Cayne, Alexander, and Lam 1994; *Banking Policy Report* 1996, p. 3; *BNA's Banking Report* 1996, p. 640).

[26] Keeton (1990), pp. 56–57, and Saunders (1985), pp. 220–21, both discuss this incentive for shifts.

Table 3

$ Millions

(1)

Parent BHC*

Assets	Liabilities and net worth
Assets 300	NW** 300

Real Estate Lending Co. (RELCO)

Assets	Liabilities and net worth
Assets 2000	Liab** 1800 NW 200

Bank

Assets	Liabilities and net worth
Assets 1000	Liab 900 NW 100

(2)

Relco bears loss

Parent BHC

Assets	Liabilities and net worth
Assets 300-125	NW 300-125

Parent's assets and net worth decline by $125M

RELCO

Assets	Liabilities and net worth
Assets 2000-125	Liab 1800 NW 200-125

RELCO suffers a $125M loan loss. Net worth falls by $125M

Bank

Assets	Liabilities and net worth
Assets 1000	Liab 900 NW 100

(3)

Loss shifted to Bank

Parent BHC

Assets	Liabilities and net worth
Assets 300-100	NW 300-100

Parent's assets and net worth decline by $100M

RELCO

Assets	Liabilities and net worth
Assets 2000	Liab 1800 NW 200

Bank

Assets	Liabilities and net worth
Assets 1000-125	Liab 900-25 NW 100-100

RELCO's $125M loss is shifted to Bank such that value of Bank's assets decline by $125M. Net worth is consumed and creditors (FDIC) bear remaining $25M loss.

* Here I focus on the parent corporation. One could focus instead on the consolidated enterprise—all assets, liabilities, and net worth on one balance sheet. The results are equivalent.
** NW is net worth, Liab is liabilities.

its true economic value well below its book capital and smaller than the value of an affiliate, therefore making the shift worthwhile.

Similarly, when a bank's true economic value sinks below zero, limited liability means that the BHC benefits from shifts of losses from the nonbank to the bank just as it benefits from shifts of income in the opposite direction. Here, the shifts simply amount to transfers from the deposit insurance fund to the

BHC's shareholders. For example, the BHC has an incentive to have the bank provide subsidized loans to affiliates or purchase services from nonaffiliates at prices higher than those that would be paid in arm's-length transactions.[27]

An earlier discussion noted a counterincentive weighing against the incentive to shift losses to defend the BHC's reputation. Here there is a similar counterincentive. Sinking the bank to take advantage of limited liability protection is likely to have serious reputation consequences for the surviving BHC. Only if the avoided losses exceed these reputation costs will the BHC choose to shift losses and cause the bank's failure.

4. CONCLUSION

This article has identified circumstances under which a BHC's shareholders will have an economic incentive to shift nonbank losses to affiliated banks. Congress and the bank regulators have reason to be concerned with any such shifts which may increase the risk of, or produce, bank failures and claims on the federal deposit insurance fund. The firewalls may provide a valuable regulatory tool for limiting shifts of nonbank losses to banks and therefore for containing BHC shareholders' economic incentive to employ such shifts.

REFERENCES

Banking Policy Report. "Cross-Guarantee Case Appealed to Supreme Court," vol. 15 (June 17, 1996).

Black, Fischer, Merton H. Miller, and Richard A. Posner. "An Approach to the Regulation of Bank Holding Companies," *Journal of Business*, vol. 51 (July 1978), pp. 379–412.

BNA's Banking Report. "Supreme Court: Retirement CD Ruling Left Standing As Supreme Court Opens New Term," vol. 67 (October 14, 1996), p. 640.

Board of Governors of the Federal Reserve System. "A Review of Restrictions on Director, Officer and Employee Interlocks, Cross-Marketing Activities, and the Purchase and Sale of Financial Assets between a Section 20 Subsidiary and an Affiliated Bank," *Federal Register*, vol. 61 (November 7, 1996a), p. 57679.

[27] Keeton (1990) discusses alternatives to firewalls that could offset the BHC's incentive to shift losses to the FDIC. These alternatives are the Fed's source-of-strength doctrine and FIRREA's requirement that a failed bank's surviving affiliate banks reimburse the FDIC (source of strength and the FIRREA provision are discussed in footnote 25). If, under the source-of-strength doctrine, BHCs were required to use their financial resources to assist troubled subsidiary banks, there would be no benefit to the BHC of shifting nonbank losses to banks. The FIRREA requirement means any nonbank loss must exceed the sum of the capital of all banks owned by a BHC, not just of one bank, before a shift can benefit the BHC.

__________. "Revenue Limit on Bank-Ineligible Activities of Subsidiaries of Bank Holding Companies Engaged in Underwriting and Dealing in Securities," *Federal Register,* vol. 61 (August 5, 1996b), p. 40643.

__________. *Bank Holding Company Supervision Manual*. Washington: Board of Governors, 1995.

__________. "Order Approving Applications to Underwrite and Deal in Certain Securities to a Limited Extent and to Engage in Combined Investment Advisory and Securities Brokerage Activities," *Federal Reserve Bulletin,* vol. 75 (March 1989), pp. 190–217.

__________. "Order Approving Applications to Engage in Limited Underwriting and Dealing in Certain Securities," *Federal Reserve Bulletin,* vol. 73 (June 1987), pp. 473–508.

__________. "A Discussion of Amendments to Section 23A of the Federal Reserve Act Proposed by the Board of Governors of the Federal Reserve System," an unpublished report prepared by the Board of Governors to accompany its legislative proposal to amend Section 23A, September 1981.

__________. "Miscellaneous Interpretation," *Federal Reserve Bulletin,* vol. 60 (October 1974), pp. 726–27.

Board of Governors v. MCorp Financial, 112 S.Ct 459 (1991). St. Paul, Minn.: West Publishing Co., 1991.

Boyd, John H., and Mark Gertler. "Are Banks Dead? Or Are the Reports Greatly Exaggerated?" Federal Reserve Bank of Minneapolis *Quarterly Review,* vol. 18 (Summer 1994), pp. 2–23.

Cayne, Howard N., Richard M. Alexander, and Brian E. J. Lam. "Federal Courts Disagree over Constitutionality of FIRREA Cross-Guarantees," *Banking Policy Report,* vol. 13 (November 7, 1994), pp. 4–6.

Comptroller of the Currency. *Annual Report,* 1977.

Cornyn, Anthony, Gerald Hanweck, Stephen Rhoades, and John Rose. "An Analysis of the Concept of Corporate Separateness in BHC Regulation from an Economic Perspective," in *Proceedings of a Conference on Bank Structure and Competition,* (Federal Reserve Bank of Chicago, May 14–16, 1986), pp. 174–212.

Federal Deposit Insurance Corporation. *Annual Report,* 1953.

Fein, Melanie L. *Securities Activities of Banks*. Englewood Cliffs, New Jersey: Aspen Publishers, Inc., 1996.

Fitzpatrick v. Federal Deposit Insurance Corp., 765 F.2d 569 (6th Cir. 1985). St. Paul, Minn.: West Publishing Co., 1985.

Gilbert, Alton R. “Do Bank Holding Companies Act As ‘Sources of Strength’ for Their Bank Subsidiaries?” Federal Reserve Bank of St. Louis *Review,* vol. 73 (January/February 1991), pp. 3–18.

__________. “A Comparison of Proposals to Restructure the U.S. Financial System,” Federal Reserve Bank of St. Louis *Review,* vol. 70 (July/August 1988), pp. 58–75.

Keeley, Michael C., and Barbara A. Bennett. “Corporate Separateness,” Federal Reserve Bank of San Francisco *Weekly Letter,* June 3, 1988.

Keeton, William R. “Bank Holding Companies, Cross-Bank Guarantees, and Source of Strength,” Federal Reserve Bank of Kansas City *Economic Review,* vol. 75 (May/June 1990), pp. 54–67.

Kelly, Edward J., III. “Conflicts of Interest: A Legal View,” in Ingo Walter, ed., *Deregulating Wall Street: Commercial Banking Penetration of the Corporate Securities Market*. New York: John Wiley and Sons, 1985a.

__________. “Legislative History of the Glass-Steagall Act,” in Ingo Walter, ed., *Deregulating Wall Street: Commercial Banking Penetration of the Corporate Securities Market*. New York: John Wiley and Sons, 1985b.

Macey, Jonathan R., and Geoffrey P. Miller. *Banking Law and Regulation*. Boston: Little, Brown and Company, 1992.

Miles, Veryl Victoria. “Banking Affiliate Regulation under Section 23A of the Federal Reserve Act,” *The Banking Law Journal,* vol. 105 (November/December 1988), pp. 476–509.

Morgan Guaranty. “An Investigation of Commercial Bank Failures,” *Memorandum,* September 1, 1983.

Perkins, Edwin J. “The Divorce of Commercial and Investment Banking: A History,” *The Banking Law Journal,* vol. 88 (June 1971), pp. 483–528.

Posner, Richard A. *Economic Analysis of Law*. Boston: Little, Brown and Company, 1977.

Rose, John T., and Samuel H. Talley. “Bank Transactions with Affiliates: The New Section 23A,” *Banking Law Journal,* vol. 100 (May/June 1983), pp. 423–38.

__________. “The Banking Affiliates Act of 1982: Amendments to Section 23A,” Board of Governors of the Federal Reserve System *Federal Reserve Bulletin,* vol. 68 (November 1982), pp. 693–99.

Saunders, Anthony. “Bank Holding Companies: Structure, Performance, and Reform,” in William S. Haraf and Rose Marie Kushmeider, eds., *Restructuring Banking and Financial Services in America*. Washington: American Enterprise Institute for Public Policy Research, 1988.

__________. "Conflicts of Interest: An Economic View," in Ingo Walter, ed., *Deregulating Wall Street: Commercial Banking Penetration of the Corporate Securities Market*. New York: John Wiley and Sons, 1985.

Seiberg, Jaret. "38 Banks Getting Head Start on Securities Business," *American Banker,* May 10, 1996.

Sinkey, Joseph F., Jr. *Problem and Failed Institutions in the Commercial Banking Industry*. Greenwich, Conn.: JAI Press Inc, 1979.

Smith, Adam. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. London: George Routledge and Sons, Limited, 1913.

Spong, Kenneth. *Banking Regulation: Its Purposes, Implementation, and Effects*, 4th ed. Kansas City: Federal Reserve Bank, 1995.

U.S. Code: Congressional and Administrative News. St. Paul, Minn.: West Publishing Company, 1982, 1966, 1956, 1955.

U.S. Congress, House. *Financial Institutions Reform, Recovery, and Enforcement Act of 1989*. House Report No. 101–222, 101 Cong. 1 Sess. Washington: Government Printing Office, 1989.

__________. *Report to Accompany H.R. 6227*. House Report, No. 609, 84 Cong. 1 Sess. Washington: Government Printing Office, 1955.

U.S. Congress, Senate. *Operation of the National and Federal Reserve Banking Systems*. Senate Report No. 77, 73 Cong. 1 Sess. Washington: Government Printing Office, 1933.

__________. *Operation of the National and Federal Reserve Banking Systems: Hearings Pursuant to S. Res. 71*. Hearing, 71 Cong. 3 Sess. Washington: Government Printing Office, 1931.

U.S. General Accounting Office. "Banks' Securities Activities: Oversight Differs Depending on Activity and Regulator." Washington: U.S. General Accounting Office, September 21, 1995.

__________. "Bank Powers: Insulating Banks from the Potential Risks of Expanded Activities." Washington: U.S. General Accounting Office, April 14, 1987.

Wall, Larry D. "Insulating Banks from Nonbank Affiliates," Federal Reserve Bank of Atlanta *Economic Review,* vol. 69 (September 1984), pp. 18–28.

Neoclassical vs. Endogenous Growth Analysis: An Overview

Bennett T. McCallum

After a long period of quiescence, growth economics has in the last decade (1986–1995) become an extremely active area of research—both theoretical and empirical.[1] To appreciate recent developments and understand associated controversies, it is necessary to place them in context, i.e., in relation to the corpus of growth theory that existed prior to this current burst of activity. This article's exposition will begin, then, by reviewing in Sections 1–4 the *neoclassical* growth model that prevailed as of 1985. Once that has been accomplished, in Section 5 we shall compare some crucial implications of the neoclassical model with empirical evidence. After tentatively concluding that the neoclassical setup is unsatisfactory in several important respects, we shall then briefly describe a family of "endogenous growth" models and consider controversies regarding these two classes of theories. Much of this exposition, which is presented in Sections 6–8, will be conducted in the context of a special-case example that permits an exact analytical solution so that explicit

The author is H. J. Heinz Professor of Economics, Carnegie Mellon University, adviser to the Research Department of the Federal Reserve Bank of Richmond, and Research Associate of the National Bureau of Economic Research. He is indebted to Michael Dotsey, Marvin Goodfriend, Robert King, Edward Nelson, Sergio Rebelo, Pierre-Daniel Sarte, and Roy Webb for helpful comments on an earlier draft. The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Notable contributions have been made by Romer (1986, 1989, 1990), Lucas (1988), Rebelo (1991), King and Rebelo (1990), Levine and Renelt (1992), Barro and Sala-i-Martin (1992, 1995), Summers and Heston (1988), and Mankiw, Romer, and Weil (1992), among others. A comprehensive and authoritative treatment of growth analysis has recently been provided by Barro and Sala-i-Martin (1995).

comparisons can be made. Finally, some overall conclusions are tentatively put forth in Section 9. These conclusions, it can be said in advance, are broadly supportive of the endogenous growth approach. Although the article contends that this approach does not strictly justify the conversion of "level effects" into "rate of growth effects," which some writers take to be the hallmark of endogenous growth theory, it finds that the quantitative predictions of such a conversion may provide good approximations to those strictly implied.

1. BASIC NEOCLASSICAL SETUP

Consider an economy populated by a large (but constant) number of separate households, each of which seeks at an arbitrary time denoted $t = 1$ to maximize

$$u(c_1) + \beta u(c_2) + \beta^2 u(c_3) + \ldots, \tag{1}$$

where c_t is the per capita consumption of a typical household member during period t and where $\beta = 1/(1+\rho)$ with $\rho > 0$ the rate of time preference. The instantaneous utility function u is assumed to be well behaved, i.e., to have the properties $u' > 0,\ u'' < 0,\ u'(0) = \infty,\ u'(\infty) = 0$. The analysis would not be appreciably altered if leisure time were included as a second argument, but to keep matters simple, leisure will not be recognized in what follows. Instead, it will be presumed that each household member inelastically supplies one unit of labor each period.

It is assumed that the number of individuals in each household grows at the rate ν; thus each period the number of members is $1+\nu$ times the number of the previous period. In light of this population growth, some analysts postulate a household utility function that weights each period's $u(c_t)$ value by the number of household members, a specification that is effected by setting $\psi = 1$ in the following more general expression:

$$u(c_1) + (1+\nu)^{\psi}\beta u(c_2) + (1+\nu)^{2\psi}\beta^2 u(c_3) + \ \ldots. \tag{1$'$}$$

With $\psi = 0$, expression (1′) reduces to (1) whereas ψ values between 0 and 1 provide intermediate assumptions about this aspect of the setup. Most of what follows will presume $\psi = 0$, but the more general formulation (1′) will be referred to occasionally.

Each household operates a production facility with input-output possibilities described by a production function $Y_t = F(K_t, N_t)$, where N_t and K_t are the household's quantities of labor and capital inputs with Y_t denoting output during t. The function F is presumed to be homogeneous of degree one so, by letting y_t and k_t denote per capita values of Y_t and K_t, we can write

$$y_t = f(k_t), \tag{2}$$

where $f(k_t) \equiv F(k_t, 1)$. It is assumed that f is well behaved (as defined above).

Letting v_t denote the per capita value of (lump-sum) government transfers (so $-v_t$ = net taxes), the household's budget constraint for period t can be written in per capita terms as

$$f(k_t) + v_t = c_t + (1+\nu)k_{t+1} - (1-\delta)k_t. \tag{3}$$

Here δ is the rate of depreciation of capital. As of time 1, then, the household chooses values of $c_1, c_2, \ldots$ and $k_2, k_3, \ldots$ to maximize (1) subject to (3) and the given value of k_1. The first-order condition necessary for optimality can easily be shown to be

$$(1+\nu)u'(c_t) = \beta u'(c_{t+1})[f'(k_{t+1}) + 1 - \delta], \tag{4}$$

and the relevant transversality condition is[2]

$$\lim_{t\to\infty} k_{t+1}\beta^{t-1}u'(c_t) = 0. \tag{5}$$

The latter provides the additional side condition needed, since only one initial condition is present, for (3) and (4) to determine a unique time path for c_t and $k_{t+1.}$ Satisfaction of conditions (3), (4), and (5) is necessary and sufficient for household optimality.[3]

To describe this economy's competitive equilibrium, we assume that all households are alike so that the behavior of each is given by (3), (4), and (5).[4] The government consumes output during t in the amount g_t (per person), the value of which is determined exogenously. For some purposes one might want to permit government borrowing, but here we assume a balanced budget. Expressing that condition in per capita terms, we have

$$g_t + v_t = 0. \tag{6}$$

For general competitive equilibrium (CE), then, the time paths of c_t, k_t, and v_t are given by (3), (4), and (6), plus the transversality condition (5). In most of what follows, it will be assumed that $g_t = v_t = 0$, in which case the CE values

[2] The role of this condition is outlined in Appendix A.

[3] Some references to proofs, in the context of a version that includes stochastic technology shocks, are given in McCallum (1989).

[4] Actually, the general equilibrium character of the analysis would be more apparent if we were to distinguish between quantities of capital supplied and demanded (per capita) by household h, writing them as $k_t(h)$ and $k_t^d(h)$. Then market clearing for period t would be represented by the condition $\Sigma k_t(h) = \Sigma k_t^d(h)$, where the summation is over all households. But with all households being treated as alike, which we do for simplicity, that condition reduces to $k_t(h) = k_t^d(h)$, so nothing is lost by failing to make the distinction. A similar conclusion is applicable to labor demand and supply, so the economy under discussion should be thought of as one with markets for capital and labor, even though these do not appear explicitly in the discussion. Also, the presence of a loan market is implicitly assumed, with one-period loans and capital serving a household equally well as stores of value.

of c_t and k_t are given by (4) and

$$f(k_t) = c_t + (1+\nu)k_{t+1} - (1-\delta)k_t, \tag{7}$$

provided that they satisfy (5).

Much interest centers on CE paths that are *steady states*, i.e., paths along which every variable grows at some constant rate.[5] It can be shown that in the present setup, with no technical progress, any steady state is characterized by stationary (i.e., constant) values of c_t and k_t.[6] (These constant values imply growth of economy-wide aggregates at the rate ν, of course.) Thus from (4) we see that the CE steady state is characterized by $f'(k) + 1 - \delta = (1+\nu)(1+\rho)$ or

$$f'(k) - \delta = \nu + \rho + \nu\rho. \tag{8}$$

This says that the net marginal product of capital is approximately (i.e., neglecting the interaction term $\nu\rho$) equal to $\nu + \rho$, a condition that should be kept in mind. If the more general utility function (1′) is adopted, the corresponding result is $f'(k) + 1 - \delta = (1+\rho)(1+\nu)^{1-\psi}$. Thus with $\psi = 1$, i.e., when household utility is $u(c)$ times household size, we have $f'(k) - \delta = \rho$.

It can be shown that, in the model at hand, the CE path approaches the CE steady state as time passes. Given an arbitrary k_1, in other words, k_t approaches the value k^* that satisfies (8) as $t \to \infty$. This result can be clearly and easily illustrated in the special case in which $u(c_t) = \log c_t$, $f(k_t) = Ak_t^\alpha$, and $\delta = 1$.[7] (Below we shall refer to these as the "LCD assumptions," L standing for log and CD standing for both Cobb-Douglas and complete depreciation.) In this case, equations (4) and (7) become

$$\frac{(1+\nu)}{c_t} = \frac{\beta\alpha A k_{t+1}^{\alpha-1}}{c_{t+1}} \tag{9}$$

and

$$Ak_t^\alpha = c_t + (1+\nu)k_{t+1}. \tag{10}$$

---

[5] Some authors use the term "balanced growth" for such paths. To me it seems preferable to use "steady state" so as to suggest a generalization of the concept of a stationary state, in which case every variable must grow at the constant rate of zero.

[6] That conclusion can be justified as follows. In (4), $u'(c_t) \equiv \lambda_t$ is an important variable. For it to grow at a constant rate, λ_{t+1}/λ_t must be constant through time. But by (4) that implies that $f'(k_{t+1})$ must be constant and so the properties of f imply that k is constant. Then we draw upon the algebraic requirement that for any three variables related as $y_t = x_t + z_t$, all three can grow at constant rates only if the rates are equal. (This can be seen by writing $y_t - y_{t-1} = x_t - x_{t-1} + z_t - z_{t-1}$ and then dividing by y_{t-1} : $(y_t - y_{t-1})/y_{t-1} = (x_t - x_{t-1})/y_{t-1} + (z_t - z_{t-1})/y_{t-1} = (x_{t-1}/y_{t-1})(x_t - x_{t-1})/x_{t-1} + (z_{t-1}/y_{t-1})(z_t - z_{t-1})/z_{t-1}$. Then if the growth rate of x, $(x_t - x_{t-1})/x_{t-1}$, exceeds that of z, the growth rate of y will increase as time passes—so the rates must be the same for x and z, and thus for y.) But then the budget constraint (3) implies, by repeated application of the foregoing requirement, that c_t, y_t, and v_t must all grow at the same rate as k_t, i.e., zero.

[7] Throughout, $\log x$ denotes the natural logarithm of x.

Since the value of k_t^α summarizes the state of the economy at time t, it is a reasonable conjecture that k_{t+1} and c_t will each be proportional to k_t^α. Substitution into (9) and (10) shows that this guess is correct and that the constants of proportionality are such that $k_{t+1} = \alpha\beta(1+\nu)^{-1}Ak_t^\alpha$ and $c_t = (1-\alpha\beta)Ak_t^\alpha$. These solutions in fact satisfy the transversality condition (TC) given by (5), so they define the CE path. The k_t solution can then be expressed in terms of the first-order linear difference equation

$$\log k_{t+1} = \log[\alpha\beta A/(1+\nu)] + \alpha \log k_t, \tag{11}$$

which can be seen to be dynamically stable since $|\alpha| < 1$. Thus $\log k_t$ converges to $(1-\alpha)^{-1}\log[\alpha\beta A/(1+\nu)]$. For reference below, we note that subtraction of $\log k_t$ from each side of (11) yields

$$\log k_{t+1} - \log k_t = (1-\alpha)[\log k^* - \log k_t], \tag{12}$$

where $k^* = [\alpha\beta A/(1+\nu)]^{1/(1-\alpha)}$, so $1-\alpha$ is in this special case a measure of the speed of convergence of k_t to k^*.

It might be thought that the complete-depreciation assumption $\delta = 1$ renders this special case unusable for practical or empirical analysis. But such a conclusion is not inevitable. What is needed for useful application, evidently, is to interpret the model's time periods as pertaining to a span of calendar time long enough to make $\delta = 1$ a plausible specification—say, 25 or 30 years. Then the parameters A, β, and ν must be interpreted in a corresponding manner. Suppose, for example, that the model's time period is 30 years in length. Thus if a value of 0.98 was believed to be appropriate for the discount factor with a period length of one year, the appropriate value for β with 30-year periods would be $\beta = (0.98)^{30} = 0.545$. Similarly, if the population growth parameter is believed to be about one percent on an annual basis, then we would have $1+\nu = (1.01)^{30} = 1.348$. Also, a realistic value for A would be about $10k^{(1-\alpha)}$, since it makes $k/y = 3/30 = 0.1$. So the LCD assumptions could apparently be considered for realistic analysis, provided that one's interest is in long-term rather than cyclical issues.[8]

2. TECHNICAL PROGRESS

Since the foregoing model approaches a steady state in which per capita values are constant over time, it may seem to be a strange framework for the purpose of *growth* analysis. But in the neoclassical tradition, growth in per capita values is provided by assuming that steady technical progress occurs, continually shifting the production frontier as time passes. With technical

[8] It is true, of course, that $\delta = 1$ implies a qualitatively different time profile for the depreciation of capital than when $\delta < 1$, since in the latter case a given stock of capital will never disappear entirely in finite time. But the usual assumption is made more for tractability than because of any evidence that it is truly representative of actual physical decay processes.

progress proceeding at the rate γ, the production function would in general be written as $Y_t = F(K_t, N_t, (1+\gamma)^t)$.[9] It transpires, however, that steady-state growth is only possible when technical progress occurs in a "labor-augmenting" fashion, i.e., when

$$Y_t = F(K_t, (1+\gamma)^t N_t).^{10} \tag{13}$$

But then with F homogeneous of degree one, we have

$$\hat{y}_t = f(\hat{k}_t), \tag{14}$$

where $y_t \equiv Y_t/(1+\gamma)^t N_t$ and $\hat{k}_t \equiv K_t/(1+\gamma)^t N_t$ are values of output and capital per "efficiency unit" of labor. Alternatively, $y_t = y_1(1+\gamma)^t f(k_t/k_1(1+\gamma)^t)$. The household's budget constraint, when expressed in terms of these variables (and with $v_t \equiv 0$), becomes

$$f(\hat{k}_t) = c_t/(1+\gamma)^t + (1+\nu)(1+\gamma)\hat{k}_{t+1} - (1-\delta)\hat{k}_t. \tag{15}$$

Maximizing (1) subject to (15) gives rise to the following first-order condition, analogous to (4):

$$(1+\nu)(1+\gamma)u'(c_t)(1+\gamma)^t = \beta u'(c_{t+1})(1+\gamma)^{t+1}[f'(\hat{k}_{t+1}) + 1 - \delta]. \tag{16}$$

In addition, we have the transversality condition

$$\lim_{t\to\infty} \hat{k}_{t+1}\beta^{t-1}u'(c_t)(1+\gamma)^t = 0. \tag{17}$$

Since there are no additional equilibrium conditions, presuming that $g_t = v_t = 0$, competitive equilibrium time paths of c_t and k_t are determined by (15) and (16), given the initial value of k, and the limiting condition (17).

Now, in order for steady growth of both c_t and $u'(c_t)$ to be possible, it will be assumed that agents' preferences are such that the function $u'(c_t)$ has a constant elasticity.[11] For reasons of symmetry, the function is usually written as

$$u(c_t) = \frac{c_t^{1-\theta} - 1}{1-\theta} \qquad \theta > 0, \tag{18}$$

---

[9] We assume that this rate γ satisfies $\gamma < \rho$. If instead we had $\gamma \geq \rho$, then the value of c_t would grow rapidly enough that in a steady state the infinite series (1) would not be convergent. Such a situation gives rise to mathematical complexities that are beyond the scope of the present exposition.

[10] For a proof that this form of technical progress is necessary for steady-state growth, see Appendix B. Of course, it is true that if the production function is Cobb-Douglas, then both Hicks-neutral and capital-augmenting technical progress are equivalent to the labor-augmenting type.

[11] Equation (16) implies that, if $\hat{k}_t$ is to be constant as it must be in a steady state, we must have $u'(c_{t+1}) = \psi u'(c_t)$ where ψ is a constant. Let $c_{t+1} = (1+\gamma)c_t$ and differentiate the previous expression with respect to c_t, obtaining $u''(c_{t+1})c_{t+1}/u'(c_{t+1}) = u''(c_t)c_t/u'(c_t)$, which implies that $u'(c)$ has the same elasticity for all values of c.

which has an elasticity of marginal utility of $-\theta$ and reduces to $u(c_t) = \log c_t$ in the special limiting case in which $\theta \to 1$.[12] Using (18), then, we rewrite (16) as

$$(1+\rho)(1+\nu)(c_t/c_{t+1})^{-\theta} = f'(\hat{k}_{t+1}) + 1 - \delta. \tag{19}$$

Finally, we define $\hat{c}_t = c_t/(1+\gamma)^t$, which implies that $c_t/c_{t+1} = \hat{c}_t/\hat{c}_{t+1}(1+\gamma)$, so we can rewrite (19) once more as

$$(1+\rho)(1+\nu)(\hat{c}_{t+1}/\hat{c}_t)^{\theta}(1+\gamma)^{\theta} = f'(\hat{k}_{t+1}) + 1 - \delta. \tag{20}$$

The latter shows that k_t will be constant in the CE steady state, and (15) then implies that the same will be true for $\hat{c}_t$. Thus we see that the per capita variables k_t, c_t, and y_t will all grow at the rate γ. Thus equation (20) shows that the (constant) value of $f'(\hat{k})$, which equals the marginal product of capital in unadjusted units, will satisfy

$$f'(\hat{k}) - \delta = (1+\rho)(1+\nu)(1+\gamma)^{\theta} - 1. \tag{21}$$

We can approximate $(1+\gamma)^{\theta}$ with $1+\gamma\theta$, assuming γ is small in relation to 1.0, so dropping cross-product terms we have the approximation

$$f'(\hat{k}) - \delta = \rho + \nu + \gamma\theta. \tag{22}$$

In the special case with $u(c_t) = \log c_t$, i.e., with $\theta = 1$, the right-hand side of (22) becomes $\rho + \nu + \gamma$. Furthermore, with the other LCD assumptions it can easily be verified that $\hat{k}_t$ behaves as k_t does in Section 2. In particular, with $\hat{y}_t = A\hat{k}_t^{\alpha}$ we have

$$\log \hat{k}_{t+1} - \log \hat{k}_t = (1-\alpha)[\log \hat{k}^* - \log \hat{k}_t], \tag{23}$$

where $\log \hat{k}^* = (1-\alpha)^{-1} \log[\alpha\beta A/(1+\nu)(1+\gamma)]$, implying that $\hat{k}_t$ approaches $\hat{k}^*$ as time passes with $1-\alpha$ being the rate of convergence.

3. OPTIMALITY

For social optimality, one would want to maximize the typical household's utility subject to the economy's overall resource constraint. But in the case in which $g_t = v_t = 0$, this constraint is exactly the same as the household's budget constraint, if each is expressed in per capita terms. So the social optimization problem becomes formally indistinguishable from the one solved by a typical household. Accordingly, the CE path will satisfy all the conditions for social optimality. This result would not hold, however, if $g_t > 0$ were financed by

[12] To find the limit as $\theta \to 1$ of $(c^{1-\theta} - 1)/(1-\theta)$, we use l'Hopital's rule to express it as the ratio of the limits of $d(c^{1-\theta} - 1)/d\theta = -c^{(1-\theta)} \log c$ and $d(1-\theta)/d\theta = -1$. Thus for $\theta \to 1$, we have $-c^0 \log c/(-1) = \log c$.

taxes that are distorting or if the model were modified so as to reflect some sort of externality.[13]

Now suppose that we ask the following question: In the model with technical progress, what $\hat{k}$ will yield the highest value of $\hat{c}$ that can be permanently sustained? In other words, among steady states that are not necessarily CE paths, which one yields the highest value of $\hat{c}$? Clearly, the budget constraint (15) implies

$$\hat{c} = f(\hat{k}) - (1+\nu)(1+\gamma)\hat{k} + (1-\delta)\hat{k} \tag{24}$$

for *any* steady state, so one can maximize $\hat{c}$ by differentiating the right-hand side with respect to $\hat{k}$ and setting the result equal to zero. We find that

$$0 = f'(\hat{k}) - (1+\nu)(1+\gamma) + (1-\delta) \tag{25}$$

is the implied condition on $\hat{k}$. Approximately, then,

$$f'(\hat{k}) - \delta = \nu + \gamma, \tag{26}$$

where the cross-product term $\nu\gamma$ has been dropped.

It will be noticed immediately that this implied "golden rule" value of $\hat{k}$, which we call $\hat{k}^+$, does not agree with the one given by (22) as the steady-state value $\hat{k}^*$ that is approached by the CE path.[14] Also, we see that, with $f' < 0, \hat{k}^+$ will normally exceed $\hat{k}^*$ since $\nu + \gamma$ will normally be smaller than $\nu + \rho + \theta\gamma$. (The latter will clearly be true under our assumption of $\gamma < \rho$ that guarantees convergence of [1].) Here the main point is that $\hat{k}^*$ has been found to be socially optimal, since it is the value approached by the CE under conditions that make CE paths satisfy all the requirements for social optimality. This fact sometimes generates confusion, since the steps leading to (26) seem to make $\hat{k}^+$ optimal from a steady-state perspective, as it gives a value of $\hat{c}$ larger than with $\hat{k}^*$. But because of time preference—i.e., $\rho > 0$ or $\beta < 1$—an economy in a steady state with $\hat{k}_t = \hat{k}^+$ could increase the value of (1) by immediately consuming slightly more than the golden rule amount, given by (24) with $\hat{k}^+$, and moving to a steady state with $\hat{k}$ somewhat smaller than $\hat{k}^+$. And so long as $\hat{k} > \hat{k}^*$, this same possibility remains open. Thus we conclude again that the optimal path beginning at any time would be given by (23), which implies that k_t will approach $\hat{k}^*$ as time passes.

[13] Suppose, for example, that $g_t > 0$ and is financed by a tax on production at the rate τ so that the (per capita) government budget constraint is $g_t = \tau f(k_t)$. Then a typical household's condition analogous to (4) becomes $(1+\upsilon)u'(c_t) = \beta u'(c_{t+1})[(1-\tau)f'(k_{t+1}) + 1 - \delta]$, but the counterpart for social optimality does not include the $(1-\tau)$ term. For a steady state with no technical progress, then, we would have $(1-\tau)f'(k) - \delta = (1+\upsilon)(1+\rho) - 1$ in CE which makes $f'(k)$ larger than optimal—that is, too little capital is accumulated (even conditional on g_t). An externality example will be considered below, in Section 6.

[14] The value $\hat{k}^*$ is frequently referred to as the "modified golden rule" value.

One way to understand the foregoing is to note that the golden rule path yields "steady-state optimality" only for the highly artificial problem of first imposing a steady-state restriction and then optimizing, instead of optimizing and then finding what conditions would be implied by an equilibrium that happens to be a steady state. The latter comes much closer to answering a relevant operational question. Or, to put the contrast in other words, k^+ is the answer to the question "what capital stock would your economy like to be miraculously given under the condition that it be *required* to maintain that value forever?"[15] By contrast, $\hat{k}^*$ is the answer to the question, "Among capital stocks that your economy would *willingly* maintain forever, which one is most desirable?"

It might be noted, incidentally, that the importance of this point regarding steady-state optimality would not be diminished by the presence of money—i.e., a transaction-facilitating medium of exchange—in the economy under consideration. Thus it remains true in monetary models that optimal steady-state paths are those found by conducting optimality analysis prior to the imposition of steady-state conditions. Failure to proceed in this manner has led to misleading conclusions or suggestions by several analysts and even mars the widely used graduate textbook of Blanchard and Fischer (1989, Chapter 4).[16]

4. HISTORY OF THOUGHT

Before continuing, let us pause to note that development of the neoclassical growth model is frequently attributed to Ramsey (1928), Cass (1965), and Koopmans (1965), with an extension to a stochastic environment provided by Brock and Mirman (1972). These papers were all concerned, however, only with the social planning problem, not with market outcomes. Recognition that the mathematical expressions could be reinterpreted so as to provide a positive theory of the behavior of a competitive market economy was first made in print—as far as I have been able to determine—by Brock (1974a). Extension to a monetary economy was accomplished by Brock (1974b).

A justly famous paper by Solow (1956) developed an analysis of growth that is in several ways closely related to the one provided by the neoclassical model. Solow's paper did not include dynamic optimizing analysis of households' saving behavior, however, but simply took the fraction of income saved to be a given constant. A contemporaneous paper by Swan (1956) developed

[15] In efficiency-adjusted per capita units, that is.

[16] Specifically, Blanchard and Fischer suggest on page 191 that the Chicago Rule optimal inflation result—i.e., that the optimal steady inflation rate equals the negative of the real rate of return on capital—depends upon the property of monetary superneutrality, and on page 181 they state that the Chicago Rule is not optimal in a monetary overlapping generations model. Both of these claims are overturned, however, by analysis that imposes steady-state restrictions after conducting a more general optimality analysis (see McCallum [1990], pp. 976–78, and 983).

a rather similar analysis in a fashion that was less mathematically explicit. Discussion of the "golden rule" condition can be found in papers by Phelps (1961) and Solow (1962).

Prior to publication of the Solow and Swan papers, considerable attention had been given to a result of Harrod (1939) and Domar (1947) to the effect that in a steady state the product of the saving/output ratio and the output/capital ratio must equal the rate of growth of capacity output. In other words, k_t and the capacity level of output $\bar{y}_t$ must grow at the same rate if $y_t/\bar{y}_t$ is to remain constant (as was taken to be necessary). Much was made of the idea that these three numbers might be determined by different aspects of economic behavior, and it was suggested that satisfaction of the condition might be unlikely to result in market economies without activist government policy. Solow (1956) cogently observed that the output/capital ratio could adjust endogenously, but—as Hahn (1987) has noted—this observation does not actually speak to the Harrod-Domar "problem." That is because Solow showed that k_t and y_t could grow at equal rates, but in doing so, he *assumed* that $y_t/\bar{y}_t$ was constant, which was actually the matter of concern to Harrod and Domar. Solow's contribution was great, nevertheless, because he (and Swan) developed something that might reasonably be called a *model*, whereas Harrod and Domar had only derived (via elementary algebra) a *condition* that needed to be satisfied for steady growth.[17]

The resurgence of growth theory that took place in the 1980s, and involved the development of endogenous growth models, arose in response to a perception that the neoclassical framework was severely inadequate for the analysis of actual growth experiences. To detail the perceived inadequacies and the subsequent response is the purpose of the next two sections.

5. WEAKNESSES OF THE NEOCLASSICAL MODEL

The evident trouble with the neoclassical growth model outlined above is that it fails to explain even the most basic facts of actual growth behavior. To a large extent, this failure stems directly from the model's prediction that output per person approaches a steady-state path along which it grows at a rate γ that is given exogenously. For this means that the rate of growth is determined outside the model and is independent of preferences, most aspects of the production function, and policy behavior. As a consequence, the model itself suggests either the same growth rate for all economies or, depending on one's interpretation, different values about which it has nothing to say. But in reality different nations have maintained different per capita growth rates over long periods of time—and these rates seem to be systematically related

[17] The Solow contribution is not a complete optimizing model, as has been mentioned, but is a model nevertheless, in the sense of a falsifiable depiction of some economic phenomena.

to various national features, e.g., to be higher in economies that devote large shares of their output to investment. These and other failings were stressed by Romer (1986, 1987, 1989) and Lucas (1988).

Of course, the neoclassical model does imply that *transitional* growth rates will differ across economies, being faster in those that have existing capital-to-effective-labor ratios relatively far below their CE steady-state values. This observation is what prevented fundamental dissatisfaction from being openly expressed before the appearance of the Romer and Lucas papers and is one of the two lines of defense recently mentioned in a lively discussion by Mankiw (1995, p. 281).[18] But transitional phenomena cannot provide a quantitative explanation of the magnitude of long-lasting growth rate differences under the standard neoclassical presumption that the production function is reasonably close to the Cobb-Douglas form with a capital elasticity of approximately one-third (roughly capital's share of national income).[19] One way to describe the problem is to consider a comparison in which one economy's per capita output increases by a factor of 2.9 relative to another's over a period of 30 years, which is the factor that would be relevant if the first economy's average growth rate exceeded the second's by about 3.6 percent per year. (This last figure is twice the standard deviation of per capita growth rates among 114 nations over the years 1960 to 1990, as reported by Barro and Sala-i-Martin [1995], p. 3, so a sizable fraction of all nation pairs have had differences exceeding that value.) Then, with a capital elasticity of one-third, the capital stock per capita would have to increase by a factor of $2.9^3 = 24.4$ relative to the second economy, if their rates of technical progress were the same. Thus the real rate of interest—i.e., the marginal product of capital—in the first economy would fall by a relative factor of $24.4^{2/3} = 8.4$. So if the two economies had similar real interest rates at the end of the 30-year period, the first economy's rate would have been 8.4 times as high as the second's at the start of the period! But of course we do not observe in actual data changes in capital/labor ratios or real interest rates that are anywhere near as large as those magnitudes, even though we observe many output growth differentials of 3.6 percent and more.[20] Some evidence that this argument is robust to production function assumptions, and a dramatic comparison involving Japan and the United States, is provided by King and Rebelo (1993).

[18] The second line of defense, that the neoclassical model may do a reasonable job of explaining cross-country differences in the *level* of income, will be discussed below.

[19] The following discussion is adapted from McCallum (1994).

[20] The interest rate portion of the foregoing argument counterfactually presumes two closed economies but so does the usual neoclassical growth analysis.

The same general type of calculation is also relevant for cross-country comparisons. The level of per capita incomes in the industrial nations of the world are easily 10 times as high as in many developing nations. With a production function of the type under discussion, this differential implies a capital per capita ratio of $10^3 = 1000$, and therefore a ratio of marginal products of capital of $1000^{-2/3} = 1/100 = 0.01$. In other words, the real rate of return to capital is predicted to be about 100 times as high in the developing nations as in those that are industrialized. But surely a differential of this magnitude would induce enormous capital flows from rich to poor countries, flows entirely unlike anything that is observed in actuality.[21]

Another perspective on the neoclassical vs. endogenous growth issue involves the question of "convergence," which has been much discussed in the literature. From equations (14) and (23) above we see that if all nations had the same taste and technology parameters, and the same population growth rate, then they should, according to the neoclassical model, have the same steady-state level of per capita income. Thus as time passes, per capita income levels in different countries should converge to a common value, with low income countries growing more rapidly than those in which beginning per capita income levels are high. Empirically, however, it is the case that growth rates over periods such as 1960 to 1985 are virtually uncorrelated with initial-year income levels. In fact, there is a small, positive coefficient in the Mankiw-Romer-Weil (1992) sample of 98 "non-oil" countries; their cross-section regression is

$$\begin{aligned}\log y_{1985} - \log y_{1960} &= \underset{(0.38)}{-0.27} + \underset{(0.050)}{0.094}\log y_{1960} \\ \bar{R}^2 = 0.03 \qquad & SE = 0.44. \end{aligned} \tag{27}$$

The neoclassical model does not actually require, however, that population growth values are equal in various countries and does not imply that taste and technology parameters must be the same. So convergence in the "unconditional" sense of the foregoing discussion is not, it can be argued, relevant to the performance of the neoclassical model. What that model does imply, according to authors including Barro and Sala-i-Martin (1992, 1995) and Mankiw, Romer, and Weil (1992), is a concept that has been termed "conditional convergence." It will be discussed below, in Section 8.

It should be noted that the foregoing discussion does not imply that the neoclassical analysis was unproductive. On the contrary, it played a major and essential role in the development of dynamic general equilibrium analysis, the basis for much of today's economic theory. It is only as a theory of growth that it is here being criticized.

[21] For more comparisons of this type and some discussion, see Lucas (1990).

6. ENDOGENOUS GROWTH MECHANISMS

In response to the various failures of the neoclassical model, Romer, Lucas, King and Rebelo, and other scholars have developed models in which steady growth can be generated endogenously—i.e., can occur without any exogenous technical progress—at rates that may depend upon taste and technology parameters and also tax policy. There are numerous variants of such models, but several important points can be developed by focusing on three basic mechanisms. Two of these, one involving a capital accumulation externality and the second relying upon the accumulation of human capital, will be discussed in this section, with the third following in Section 7.

Let us consider first the *externality* model. For its presentation we will modify the setup of Section 1, in which there is no exogenous technical progress. There is, however, an externality in production so that the typical household's per capita production function is

$$y_t = f(k_t, \bar{k}_t) \qquad f_2 > 0_1,\ f_{22} < 0, \tag{28}$$

where $\bar{k}_t$ is the *economy-wide average* capital stock per person. Quoting Romer (1989, p. 90), the "rationale for this formulation is based on the public good character of knowledge. Suppose that new physical capital and new knowledge or inventions are produced in fixed proportions so that [$\bar{k}_t$] is an index not only of the aggregate stock of physical capital but also of the aggregate stock of public knowledge that any firm can copy and take advantage of." But each firm or household is small, so it views $\bar{k}_t$ as given when making its choice of k_{t+1} and other decision variables.

So as to highlight the effect of the resulting externality, suppose that the production function is Cobb-Douglas,

$$y_t = Ak_t^{\alpha}\bar{k}_t^{\eta}, \tag{28$'$}$$

and that the other LCD assumptions hold as well (i.e., $u(c_t) = \log c_t$ and $\delta = 1$). Also, let $g_t = v_t = 0$. Then the household's budget constraint is

$$Ak_t^{\alpha}\bar{k}_t^{\eta} = c_t + (1+\nu)k_{t+1} \tag{29}$$

and its first-order optimality condition is

$$\frac{(1+\nu)}{c_t} = \frac{\beta\alpha Ak_{t+1}^{\alpha-1}\ \bar{k}_{t+1}^{\eta}}{c_{t+1}}. \tag{30}$$

In addition, for general competitive equilibrium the following condition must be satisfied, since households are alike:

$$k_t = \bar{k}_t. \tag{31}$$

Given these relations,[22] it is a reasonable conjecture that in a CE both k_{t+1} and c_t will be proportional to $k_t^{\alpha+\eta}$. Substitution into (29) and (30), using (31), shows this guess to be correct and that the resulting expression for k_{t+1} is

$$k_{t+1} = \alpha\beta(1+\nu)^{-1}Ak_t^{\alpha+\eta}. \tag{32}$$

There are two interesting points relating to this solution. First, with $\eta > 0$ the CE path is not socially optimal. For social optimality, the problem is to maximize (1) subject not to (29), but to (29) with (31) imposed. In that case, the equation comparable to (32) that results is

$$k_{t+1} = (\alpha+\eta)\beta(1+\nu)^{-1}\ Ak_t^{\alpha+\eta}. \tag{32$'$}$$

Clearly, if $\alpha + \eta < 1$, then (32) implies that k_t approaches a constant value, but it is one that is smaller than the steady-state value implied by (32′)—an outcome that reflects the failure of individuals to take account of their own actions' effect on the economy-wide state of knowledge. Second, if by chance it happened that $\alpha+\eta = 1$, then k_t would grow forever at a constant rate equal to $\alpha\beta(1+\nu)^{-1}A-1$.[23] Thus, it is possible, within this framework that excludes exogenous technical progress, for steady-state growth to be generated, in which case its rate will be dependent upon α, β, ν, and A. Admittedly, the case with $\alpha + \eta = 1$ exactly might be regarded as rather unlikely to prevail. That issue will be taken up below.

Now let us consider the second of the two basic endogenous growth mechanisms, this one involving the accumulation of human capital—in the sense of labor-force skills that can be enhanced by the application of valuable resources. One simple way to represent this phenomenon is to specify that physical output is accumulated according to

$$Ak_t^{\alpha}(h_t n_t)^{1-\alpha} = c_t + (1+\nu)k_{t+1}, \tag{33}$$

where n_t is the fraction of the typical household's work time that is allocated to goods production and h_t is a measure of human capital—i.e., workplace skills—of a typical household member at time t.[24] These skills are produced by devoting the fraction $1-n_t$ of working time to human capital accumulation. In general, physical capital would also be an important input to this process, but for simplicity let us initially assume that the accumulation of productive skills obeys the law of motion

$$h_{t+1} - h_t = B(1-n_t)h_t - \delta_h h_t, \tag{34}$$

[22] And also a transversality condition.

[23] In this case the transversality condition requires $\lim(1+\nu)k_{t+1}\beta^{t-1}/c_t = 0$. With k_{t+1} given by (32) and $c_t = (1-\alpha\beta)Ak_t^{\alpha+\eta}$, the relevant expression is $\beta^{t-1}\alpha\beta/(1-\alpha\beta)$, which does indeed approach zero as $t \rightarrow \infty$.

[24] In (33), the h_t human capital measure enters in the same way that the labor-augmenting technical progress term does in the neoclassical setup. So it can be seen that constant growth will occur if the overall model is such that h_t grows (endogenously) at a constant rate.

where the final term reflects depreciation of skills that occurs as time passes. In this expression, and for the rest of this example, we let $\nu = 0$.

Maximization of (1) subject to constraints (33) and (34) gives rise to the following first-order conditions:

$$c_t^{-1} = \beta c_{t+1}^{-1} \alpha A k_{t+1}^{\alpha-1}(h_{t+1}n_{t+1})^{1-\alpha} \tag{35a}$$

$$c_t^{-1} A k_t^{\alpha} h_t^{1-\alpha}(1-\alpha) n_t^{-\alpha} = \mu_t B h_t \tag{35b}$$

$$\mu_t = \beta\mu_{t+1}[B(1-n_{t+1}) + 1 - \delta_h] + \beta c_{t+1}^{-1} A k_{t+1}^{\alpha} n_{t+1}^{1-\alpha}(1-\alpha)h_{t+1}^{-\alpha}. \tag{35c}$$

Here μ_t is the shadow price of human capital, i.e., the Lagrange multiplier attached to (34). With $g_t = v_t = 0$, the CE is given by the five equations (33), (34), and (35a)–(35c),[25] which determine time paths for c_t, k_t, h_t, n_t, and μ_t. Since (33) and (34) are the same from the private and social perspectives, there is no departure from social optimality implied by the CE.[26]

Now consider the possibility of steady-state growth in this system. Since n_t is limited to the interval [0,1], it must be constant in any steady state. If its value is n, then (34) shows that h_t will grow at the steady rate $B(1-n) - \delta_h$, which we now denote as ξ. Then (35a) implies, since c_{t+1}/c_t must be constant, that k_t must also grow at the rate ξ—and by (33) the same must be true for c_t. Finally, (35b) shows that $1/\mu_t$ must grow like c_t—and these conclusions are consistent with (35c) having each term grow at the same rate. To find out what this growth rate will be, we can equate μ_t from (35b) and (35c), using $\mu_{t+1} = \mu_t/(1+\xi)$ in the latter, and after some tedious simplification find that

$$\rho(1+\xi) = Bn. \tag{36}$$

Since also $\xi = B(1-n) - \delta_h$, we can solve for

$$n = \frac{\rho(1 + B - \delta_h)}{(1+\rho)B} \tag{37}$$

in terms of basic parameters of the problem. Then ξ is found easily from expression (36).

An important property of (37) to be noted is that the steady-state value of n increases with ρ. Thus ξ, the growth rate, decreases with ρ, the rate of time preference. In other words, the more impatience is exhibited by the economy's individuals, the lower will be the steady-state growth rate. This is precisely the sort of result that some analysts have found highly plausible but is not generated by the neoclassical model. If $\nu \neq 0$ is assumed, moreover, the growth rate is negatively related to ν.

---

[25] Plus a pair of transversality conditions. The present model, it should be said, is the first of two in Lucas (1988), but here it is given without an externality, and it is very similar to one developed much earlier, by Uzawa (1965).

[26] This last statement presumes the absence of government spending and distortionary taxes.

An obvious objection to the model based on (33) and (34) is that production of h_t should be specified as dependent on the use of capital—i.e., physical goods—in that process. That extension has been studied by Rebelo (1991), who uses the following in place of (34):

$$h_{t+1} - h_t = B(m_t k_t)^a [h_t(1 - n_t)]^{1-a} - \delta_h h_t. \tag{38}$$

Here m_t denotes the fraction of the capital stock that is devoted to production of human capital, so $(1 - m_t)k_t$ replaces k_t in (33) in this model.[27] Rebelo finds that the same conclusions involving steady growth and its dependence upon ρ hold with this extension. Furthermore, if production of physical output is taxed, say at the rate τ, then the steady-state growth rate will be negatively related to τ.

Of the two mechanisms considered, knowledge externalities and human capital, it is not obvious which is the more plausible as a source of major quantitative departure from the neoclassical model. But there is no reason to consider them on an either-or basis; both could be relevant simultaneously. Indeed, the Lucas (1988) model, of which our (33) and (34) are a special case, posits human capital accumulation as in (34) together with a production function for physical output in which there is an externality involving average economy-wide human—rather than physical—capital.

In what follows it will be useful to have at hand the full dynamic, period-by-period solution for a representative endogenous growth model. It is possible to derive such a solution for the Lucas model, even with the human capital production externality included, provided that we use the LCD version, which in this case requires that human capital fully depreciates in one period. Accordingly, let us now modify the model of equations (33), (34), and (35) by using $Ak_t^\alpha (h_t n_t)^{1-\alpha} \bar{h}_t^\eta$ as the production function and setting $\delta_h = 1$. Also, we shall permit population growth again, which implies that h_{t+1} in (34) and μ_t in (35c) are multiplied by $(1 + \nu)$. Then the household's optimality conditions, other than the transversality conditions, can be written as follows:

$$Ak_t^\alpha (h_t n_t)^{1-\alpha} \bar{h}_t^\eta = c_t + (1 + \nu)k_{t+1} \tag{39a}$$

$$(1 + \nu)h_{t+1} = B(1 - n_t)h_t \tag{39b}$$

$$(1 + \nu)c_{t+1} = c_t \beta \alpha A k_{t+1}^{\alpha-1} (h_{t+1} n_{t+1})^{1-\alpha} \bar{h}_{t+1}^\eta \tag{39c}$$

$$c_t^{-1} A k_t^\alpha h_t^{1-\alpha} (1 - \alpha) n_t^{-\alpha} \bar{h}_t^\eta = \mu_t B h_t \tag{39d}$$

$$(1 + \nu)\mu_t = \beta \mu_{t+1}[B - (1 - n_t)] + \beta c_{t+1}^{-1} A k_{t+1}^\alpha n_{t+1}^{1-\alpha} (1 - \alpha) h_{t+1}^{-\alpha} \bar{h}_{t+1}^\eta. \tag{39e}$$

[27] Rebelo also includes variable leisure in his setup.

In competitive equilibrium we will also have $h_t = \bar{h}_t$, so in what follows we assume that condition to hold. To solve these equations for $c_t, k_{t+1}, h_{t+1}, n_t$, and μ_t, we proceed by guessing—in analogy with the method of Section 2—that those five variables are determined in response to the state variables k_t and h_t by expressions of the form

$$c_t = \phi_{10} k_t^{\phi_{11}} h_t^{\phi_{12}} \tag{40a}$$

$$k_{t+1} = \phi_{20} k_t^{\phi_{21}} h_t^{\phi_{22}} \tag{40b}$$

$$h_{t+1} = \phi_{30} k_t^{\phi_{31}} h_t^{\phi_{32}} \tag{40c}$$

$$n_t = \phi_{40} k_t^{\phi_{41}} h_t^{\phi_{42}} \tag{40d}$$

$$\mu_t = \phi_{50} k_t^{\phi_{51}} h_t^{\phi_{52}}. \tag{40e}$$

If we can determine the implied values of the ϕ's, we will have substantiated this guess.

We begin by substituting (40c) and (40d) into (39b), obtaining

$$(1+\nu)\phi_{30} k_t^{\phi_{31}} h_t^{\phi_{32}} = Bh_t - B\phi_{40} k_t^{\phi_{41}} h_t^{\phi_{42}} h_t. \tag{41}$$

But then for (40) to be valid for all values of k_t and h_t, it must be that $\phi_{31} = \phi_{41} = \phi_{42} = 0$ and $\phi_{32} = 1$. Continuing in this manner of reasoning,[28] we end up with various sensible-looking results such as that h_t grows steadily at the rate $[B\beta/(1+\nu)] - 1$, the fraction of physical output saved is $\alpha\beta$, and especially that k_t evolves as

$$k_{t+1} = \frac{\alpha B(1-\beta)^{1-\alpha}}{(1+\nu)} A k_t^{\alpha} h_t^{1-\alpha+\eta}. \tag{42}$$

We shall make use of this last solution expression in Section 8.

An interesting and influential variant results when we again suppress the externality, by setting $\eta = 0$, but assume that human capital is produced by a production function of type (38) but with $a = \alpha$, i.e., with the same parameters as pertain to production of consumption (and physical capital) output. With log utility and Cobb-Douglas production functions, m_t and n_t will be constant over time; and with the production functions the same as well, the relative price of a unit of human capital in terms of output will be 1.0. Thus the sum of the two outputs is of the form (const.) $k_t^{\alpha} h_t^{1-\alpha}$ = (const.) $k_t(h_t/k_t)^{1-\alpha}$. But in this

[28] Specifically, we find that $n_t = \phi_{40} \equiv n$ and $h_{t+1} = \phi_{30} h_t$, with $\phi_{30} = B(1-n)/(1+\nu)$ and n yet to be determined. Next, we substitute into (39a) and in the same way find that $\phi_{11} = \phi_{21} = \alpha$ and $\phi_{12} = \phi_{22} = 1 - \alpha + \eta$. Also, substitution of (40e) and $c_t = \phi_{10} k_t^{\alpha} h_t^{1-\alpha+\eta}$ into (39d) yields $\phi_{51} = 0$ and $\phi_{52} = -1$. Finding the ϕ_{j0} values is a bit more difficult. But the equations that result when the k_t and h_t terms are canceled out of (39) with (40) inserted imply that $n = \phi_{40} = 1 - \beta$ and that $\phi_{10} = A(1-\alpha\beta)(1-\beta)^{1-\alpha}$, $\phi_{20} = A\alpha\beta(1-\beta)^{1-\alpha}/(1+\nu)$, $\phi_{30} = B\beta/(1+\nu)$, and $\phi_{50} = (1-\alpha)/B(1-\alpha\beta)(1-\beta)$.

special case it is also true that k_t/h_t is constant, so the foregoing expression reduces to a constant times k_t, often written as $y_t = Ak_t$. Hence, this is one case of the so-called "*AK*" model, which from a growth perspective is similar to an extreme special case of the neoclassical model—one in which the capital elasticity parameter α equals one. In this case, k_t and therefore output per person grows without limit at a constant rate, even with no technological progress, as inspection of equation (11) shows clearly. Furthermore, even if a and α differ so that k_t/h_t varies from period to period, the model works as indicated from a steady-state growth perspective. Consequently, the *AK* model—which may also be rationalized in other ways—has played a prominent role in the discussion of endogenous growth possibilities. We shall refer to it again shortly.

7. ISSUES CONCERNING ENDOGENOUS GROWTH ANALYSIS

Do models of the type outlined in Section 6 make more sense than the neoclassical construct that they were designed to replace? Clearly they have the virtue of at least attempting to explain growth endogenously, but are these attempts logically satisfactory and empirically plausible? In this writer's view, there are some highly attractive features of the models discussed above, including the possibility of knowledge externalities and the recognition that progress in terms of workforce skills relies in large part upon the allocation of resources to the production of such skills. But there are apparently two logical difficulties with these models that need to be considered before conclusions can be drawn.

The first of these difficulties is that in the Lucas or Lucas-Rebelo model, never-ending growth requires never-ending increases in human capital h_t, our measure of the productive skills of a typical worker. But for such a variable, never-ending growth is implausible because the skills in question are ones possessed by individual human beings and so are not automatically passed on to workers in succeeding generations. The son of a skilled craftsman is not born with dexterity and judgment but must start over again in developing them—again expending resources to do so—and has only a finite lifetime in which to do so. In this regard human capital is different from the stock of *knowledge,* which is possessed by society in general and is passed on from generation to generation, in the sense that it is available to those who wish to draw upon it.

Thus it is some form of knowledge, not human capital, that can plausibly provide the basis for never-ending growth.[29] But the development of knowledge also requires the expenditure of resources, so the question that arises is why rational private agents would devote resources to its development when the product will be possessed by society in general, rather than by themselves. A

[29] This point is stated clearly by Grossman and Helpman (1994, p. 35).

response is suggested, however, by consideration of our existing patent system. Formally, the answer in the literature is that the development of "designs" or "blueprints" that are privately profitable can have the by-product of adding to the stock of accessible productive knowledge. This type of process can continue without limit, so it can serve as the basis for never-ending growth. Several models expressing this notion have been developed;[30] let us briefly consider the most prominent of them, due to Romer (1990).

In Romer's (1990) setup the production function for (consumable) output can be written as

$$y_t = n_t^{1-\alpha} \sum_{i=1} x_{it}^{\alpha}, \tag{43}$$

where n_t is again the fraction of labor time devoted to the production of consumables and where x_{it} is the quantity of an intermediate good of type i used in period t.[31,32] The summation index i ranges from 1 to ∞ but in any period x_{it} will be zero for $i > A_t$, where A_t indicates the number of distinct intermediate goods in use. The technology for producing each intermediate good requires that ζ units of "capital" k_t must be used in the production of a unit of x_i, where capital is simply consumable output that is not consumed. Thus, if because of symmetry there is a common value $\bar{x}_t$ of x_{it} for those intermediate goods that are produced in t, we have

$$k_t = \zeta \bar{x}_t A_t. \tag{44}$$

But also $\Sigma x_{it}^{\alpha} = A_t \bar{x}_t^{\alpha}$, so substitution into (43) yields

$$y_t = n_t^{1-\alpha} A_t (k_t/\zeta A_t)^{\alpha} = (1/\zeta) n_t^{1-\alpha} k_t^{\alpha} A_t^{1-\alpha}. \tag{45}$$

From the latter it is clear that steady growth of consumable output will be possible if A_t grows exponentially without bound.

In addition to requiring ζ units of capital per unit produced, each intermediate good requires the use of one design. Designs, like output and intermediate goods, require resources in their production. Let $1 - n_t$ be the fraction of labor time devoted to the production of designs, an activity that will be called "research." Romer (1990) assumes that the research process is such that the creation of new designs by an individual is proportional to $(1 - n_t)A_t$, where A_t is the total number of intermediate good designs, not the per-person value. Thus

---

[30] Authors include Aghion and Howitt (1992), Grossman and Helpman (1991), King and Levine (1993), and Goodfriend and McDermott (1995).

[31] Romer's (1990) presentation pretends that there is a single price-taking producer of final output of consumables. Consequently, we shall not at this point distinguish between per-person and aggregative magnitudes.

[32] Actually, Romer (1990) has all labor allocated to the production of consumables and also includes a human capital measure, with some human capital used in the production of designs. Our specification is notationally simpler and basically equivalent.

Romer assumes that designs are non-rival in the research process, the activities of each researcher being enhanced by the entire stock of design knowledge accumulated to date. The evolution over time of A_t will therefore conform to

$$A_{t+1} - A_t = \sigma N(1 - n_t)A_t, \tag{46}$$

where σ is the constant of proportionality and N is the number of researchers each devoting $(1 - n_t)$ units of labor to research activity in t. Here the crucial allocation problem is the determination of $1 - n_t$, the fraction of time devoted to research instead of consumable output. In Romer's setup, this allocation depends upon the derived demand for research, which itself stems from the usefulness of intermediate goods in the production of output and the necessity of designs for these goods. Thus the evolution of A_t is determined by the optimizing choices of private individual agents (as well as technology). But in a steady state, which is shown to exist by Romer's careful analysis, n_t will be constant over time and A_t will grow at a constant rate, as indicated by (46).[33] Thus never-ending growth is generated in this model via endogenously rationalized, never-ending accumulation of knowledge.[34]

The second logical difficulty of the endogenous growth approach is the assumption of precisely constant returns to scale in the crucial production process. In the Lucas-Rebelo model, for instance, the sum of the exponents on physical and human capital in (33) and (38) must equal 1.00 exactly for steady-state growth to be implied; if this sum equals 0.99 instead, then the economy will approach a steady state in which there is no growth in the per capita quantities. Similarly, in the externality model $\alpha + \eta$ must equal 1.00 exactly in (28′) for steady growth to occur—this can be seen clearly in (32). And in the Romer (1990) model, the exponent on A_t on the right-hand side of (46) must be exactly 1.00. Consequently, the dramatically different properties of these models, as compared with the neoclassical construct, require very special parameter values that obtain only on measure-zero subsets of the relevant parameter spaces.[35] That must be regarded as implying that the endogenous growth approach does not actually generate steady, everlasting growth in the absence of exogenous technical progress.[36]

---

[33] Romer (1990) emphasizes the monopoly power possessed by each design creator, market power that gives the individual an incentive to devote resources to the research activity (in our exposition, by hiring labor). It should be noted that this is the rather benign type of monopoly power that is granted by patent systems.

[34] Rivera-Batiz and Romer (1991) argue that an important application of this analysis is in the international context, where it implies that growth is fostered by economic integration and trade liberalization (for both goods and ideas).

[35] This conclusion does not seem to be inconsistent with the discussion of Romer (1994, pp. 17–18), despite the difference in tone.

[36] Some analysts argue that all production functions must, as a matter of logic, have input coefficients summing to precisely 1. But even if one ignores the presence of land, which is probably of some importance, this argument misses the issue, which is whether the coefficients

Nevertheless, the endogenous growth approach has been highly productive, because with returns to scale reasonably *close* to 1.00 in (33) and (34), the model will have very slow transition dynamics. The speed of convergence of k_t to k_t^*, given in (12) as $1 - \alpha$ for the LCD neoclassical model, will be more nearly equal to $1 - (\alpha_1 + \alpha_2)$, where $\alpha_1 + \alpha_2$ is the sum of capital and human capital coefficients. Therefore, if the human capital coefficient arises as in (33), via the effect of skill-adjusted labor, one would expect $1 - (\alpha_1 + \alpha_2)$ to be close to zero and convergence to be very slow. But with very slow transition dynamics, growth rate differences due to transitional movements toward the CE steady state will be prolonged—so observed growth rate differences might be sustained over very long time periods. So even if the approach of the endogenous-growth proponents fails to explain never-ending steady-state growth, it could plausibly explain many features of the empirical data and potentially provide the basis for useful policy analysis.

In this regard, it is notable that equation (11) indicates that there is no discontinuity involving the distinction between neoclassical models with α close to 1.0 and endogenous-growth *AK* models with $\alpha = 1.0$ exactly. Specifically, if at some point in time the efficiency parameter *A* were changed by the amount Δ (say), then the effect on $\log k_t$ after *T* periods will be $T \log \Delta$ according to the *AK* model or $\log \Delta(1-\alpha^T)/(1-\alpha)$ according to the neoclassical model. For α values close to 1.0, these magnitudes are similar (and are equal in the limit as $\alpha \rightarrow 1.0$). With $\alpha = 0.98$ for example, we have $(1 - 0.98^5)/(1 - 0.98) = 4.8$ when $T = 5$ and $(1 - 0.98^{20})/(1 - 0.98) = 16.6$ when $T = 20$. So the response of capital and output to changes in *A*—or another variable that affects the steady-state value of k_t—will be reasonably similar whether the counterpart of α is 0.98 or 1.00. Since time periods in our formulation are about 25 to 30 years in duration, the similarity holds for substantial spans of time.

Of course, strictly speaking, this sort of weakened version of the approach does not result in the conversion of “level effects” into “rate-of-growth effects” that some writers take to be the hallmark of endogenous growth analysis. But the difference is not too great, quantitatively. Furthermore, while that conclusion implies a less dramatic difference between neoclassical and endogenous growth models, it also rescues the latter from evidence suggesting apparent empirical rejections. For example, Jones (1995) points out that the U.S. growth rate has not risen over the last century despite increases in some variables (e.g., investment share, R&D share) that would bring about rate-of-growth effects in the standard endogenous growth models with 1.00 values for the relevant parameters.

on k_t and h_t in (33) sum to 1, not the coefficients on k_t and n_t. In other words, with h_t generated by (34) or (38), the issue is whether effective labor is $h_t n_t$ or n_t multiplied by some nonlinear function of h_t.

It has been argued by Mankiw, Romer, and Weil (1992) and Mankiw (1995) that, although it has some weaknesses, the neoclassical model's empirical performance is much better than is suggested by the discussion of its endogenous growth critics or that of Section 5 above.[37] In particular, the neoclassical model is fairly successful in explaining cross-country differences in income *levels* and is even more successful when the role of human capital is taken into account.[38] To understand this point, let us return to the LCD version of the neoclassical model with technical progress. From the definition of k_t and the relation $\hat{k}_t^* = [\alpha\beta A/(1+\nu)(1+\gamma)]^{1/(1-\alpha)}$, we see that the steady-state value of k_t at time t will be

$$k_t = k_0(1+\gamma)^t[\alpha\beta A/(1+\nu)(1+\gamma)]^{1/(1-\alpha)}. \tag{47}$$

Thus for steady-state $\log y_t$ we have, using $\log(1+\gamma) \doteq \gamma$,

$$\log y_t = const + \gamma t + [\alpha/(1-\alpha)]\log[\alpha\beta A/(1+\nu)(1+\gamma)]. \tag{48}$$

For any given value of t, accordingly, $\log y_t$ will be larger the larger is β and the smaller are ν and γ. In the special model at hand, it happens that $\alpha\beta$ equals the fraction of income s that is saved each period. Thus it accords with the Mankiw, Romer, Weil estimation of an equation analogous to (48) on various cross-section samples of national economies with data averaged over the period 1960 to 1985.[39] They assume the same γ for all nations and are therefore able to incorporate γt into the constant term.[40] They obtain estimates with $\log s$ and $\log[(1+\nu)(1+\gamma)]$ entered separately and test the hypothesis that the slope coefficients are equal in magnitude and opposite in sign. The striking result of this exercise is that for their sample of 98 non-oil nations, the variables $\log s$ and $\log(1+\nu)(1+\gamma)$ have a considerable amount of explanatory power for $\log y_t$, the adjusted R^2 value being 0.59. Furthermore, the slope coefficient hypothesis mentioned above cannot be rejected at conventional significance levels. The one serious flaw acknowledged by Mankiw, Romer, and Weil is that the implied value of α is about 0.6, much larger than the one-third value that is usually presumed (and that matches the capital share of income). But this failure can be largely overcome, they demonstrate, by including additional variables designed to proxy for the level of human capital or labor-force skill in

[37] Actually, these authors are concerned with the Solow model, i.e., the special case of the neoclassical model in which the saving rate is given exogenously. But that difference is unimportant for the issues at hand.

[38] Account is taken in a different way than in our discussion surrounding equation (34), however, since human capital enters the production function as an additional input rather than as an efficiency term attached to labor input.

[39] Of course they do not use the $\delta = 1$ assumption that permits us to derive the $s = \alpha\beta$ result.

[40] They treat cross-country differences in $\log A$ as a component of the regression's disturbance term. This is not innocuous, as will be seen momentarily.

the various countries. Thus they conclude that "the Solow model is consistent with the international evidence if one acknowledges the importance of human as well as physical capital" (1992, p. 433).[41]

This argument is ingenious and the finding is interesting, but the suggestion that it serves to rescue the neoclassical model from its critics seems inappropriate.[42] For the model was designed to provide understanding about *growth,* not about international differences in income levels. In support of this last assertion, it may be noted that there is no mention of using the model for the latter purpose in Solow (1956, 1970, 1994), or Meade (1962), or Hahn (1987). That use seems to have been discovered by Mankiw, Romer, and Weil (1992).[43]

In this regard, another objection to the Mankiw, Romer, and Weil analysis has been expressed by Grossman and Helpman (1994). As mentioned above, that analysis assumes that γ, the rate of technological progress, is the same for all countries in their cross-section regressions—which thereby pushes γt into the constant term in expressions like (48). But, as Grossman and Helpman (1994, p. 29) say, "if technological progress [actually] varies by country, and [variation in γt] is treated as part of the unobserved error term, then ordinary least squares estimates of the . . . equation will be biased when investment-GDP ratios are correlated with country-specific productivity growth. In particular, if investment rates are high where productivity grows fast, the coefficient on the investment [or saving] variable will pick up . . . part of the variation due to their different experiences with technological progress. . . . [Furthermore,] an economist would certainly expect investment to be highest where capital productivity is growing the fastest." Thus the Mankiw, Romer, and Weil (1992) estimate of the effect of the saving/investment variable is overstated and the slope-coefficient test is consequently biased. Whether this bias is large quantitatively has not yet been established, but in any event it pertains only to the neoclassical model's role of explaining cross-section income levels, which seems rather incidental.[44]

[41] It should be said, however, that the Mankiw, Romer, and Weil (1992) measure of human capital leaves much to be desired. In particular, they use an estimate of the fraction of the working age population that is currently enrolled in secondary school. A measure of the fraction of the current working age population that attended (in the past) secondary school would be much more appropriate.

[42] Mankiw (1995, p. 423) says that from the standpoint of enhancing understanding, "the neoclassical model is still the most useful theory of growth that we have."

[43] A word of explanation is needed here since Denison (1967) and others have in fact looked at cross-country differences in income levels. The point is that they have done so in a way that relies on transitional differences, whereas the Mankiw, Romer, and Weil equation (48) pertains to steady-state differences.

[44] Very recently, Islam (1995) has implemented a panel-data approach to conditional convergence regressions, thereby permitting some production function parameters to differ across countries. His approach retained the assumption that γ is the same for all countries, however, so it does not address the specific criticism stressed by Grossman and Helpman.

8. CONDITIONAL CONVERGENCE

Because conditional convergence has figured prominently in the literature's controversies, it should be useful to describe—as promised above—this concept. From equation (23) we have $\log \hat{k}_{t+1} = \alpha \log \hat{k}_t + (1 - \alpha) \log \hat{k}^*$, which, since $\log \hat{y}_t = \log A + \alpha \log \hat{k}_t$, implies

$$\log \hat{y}_{t+1} = \alpha \log \hat{y}_t + (1 - \alpha) \log \hat{y}^*. \tag{49}$$

Iteration then shows that

$$\log \hat{y}_{t+j} = \alpha^j \log \hat{y}_t + (1 - \alpha^j) \log \hat{y}^*, \tag{50}$$

so that we have

$$\log y_{t+j} - \log y_t = (1 - \alpha^j)[\log \hat{y}^* - \log y_t] + j \log(1 + \gamma). \tag{51}$$

But $\log \hat{y}^* = \log A + \log[\alpha\beta A/(1 + \nu)(1 + \gamma)]$. Thus in a cross section of economies one needs to take account of potential cross-economy differences in taste, technology, and population-growth parameters even if it is assumed that γ is the same everywhere. But with proxies for these included in a regression with $(1/j)(\log y_{t+j} - \log y_t)$ on the left and $\log y_t$ on the right-hand side, the coefficient on the latter is predicted by this special case of the neoclassical model to be $-(1 - \alpha^j)/j$. Some simple endogenous growth models such as (29) and (30) with $\alpha + \eta = 1$ suggest, by contrast, that the coefficient on $\log y_t$ should be zero. So a significant negative coefficient would constitute evidence against these rudimentary specifications. Other two-sector versions such as (33)–(38) feature transitional dynamic adjustments, however, that are not ruled out by findings of conditional convergence. That has been established by Mulligan and Sala-i-Martin (1993) and will be demonstrated below for our version of the Lucas model. Thus the fact that most existing studies of the type under discussion do find significant negative coefficients does not discriminate between endogenous and neoclassical specifications.

It should be mentioned explicitly that the foregoing exposition makes use of the LCD assumptions, which lead to the conclusion that the coefficient on $\log y_t$ is a function only of the capital-elasticity parameter α. More generally, without those assumptions this coefficient will depend also on other parameters including ν, γ, and the rate of depreciation—at least in the standard approximation that is typically used in the literature (see Barro and Sala-i-Martin [1995], p. 81, or Mankiw [1995], p. 310).

To see that conditional convergence is a property of the LCD version of the Lucas model, as stated above, rewrite the solution equation (42) as follows:

$$\begin{aligned} \log k_{t+1} - \log k_t = (1 - \alpha)[\log h_t - \log k_t] + \eta \log h_t \\ + \log[\alpha\beta(1 - \beta)^{1-\alpha} A/(1 + \nu)]. \end{aligned} \tag{52}$$

This shows clearly that conditional convergence holds for $\log k_t$, i.e., that $\log k_t$ will have a negative slope coefficient in a regression with $\log k_{t+1} - \log k_t$ as the dependent variable, when $\log h_t, \log s_t, \log(1+\nu)$, and $\log(1-\beta)$ are included as regressors. Furthermore, the same is true for $\log y_t$, as can be seen by using the production function to eliminate k_t and k_{t+1} in favor of y_t and y_{t+1}.

Quite recently, an interesting implication of the literature's empirical findings on conditional convergence was pointed out by Cho and Graham (1996). The basic finding is that in cross-section regressions with large samples of heterogeneous countries, estimates of the slope coefficient b_1 in relations of the form

$$\log y_{t+j} - \log y_t = b_0 + b_1[\log y_t^* - \log y_t] \tag{53}$$

are positive, when expressions for y_t^* are ones suggested by the neoclassical model. But one of the main reasons that the conditional convergence formulation was invented is that in such samples of countries one frequently obtains positive estimates of b_3 in regressions of the form

$$\log y_{t+j} - \log y_t = b_2 + b_3 \log y_t. \tag{54}$$

Equating the right-hand sides of (53) and (54), however, we see that

$$b_0 + b_1[\log y_t^* - \log y_t] = b_2 + b_3 \log y_t, \tag{55}$$

plus regression residuals. But with $b_1 > 0$ and $b_3 > 0$, (55) indicates that output is smaller in relation to its steady-state value (of the current period) for high-income countries than for low-income countries. In other words, if low-income countries have less capital than in the CE steady state, then they are relatively closer to their steady-state positions than are rich nations and, in that sense, have less economic development yet to be accomplished, i.e., negative catching up! Admittedly, estimates of b_3 are quite unrealiable and often insignificantly positive. But suppose, then, that we take b_3 to be essentially zero. Then (55) suggests that low-income countries are on average neither closer to (proportionately) nor farther from their steady-state positions than are rich countries.

9. CONCLUSION

Let us conclude with a brief summary of the arguments developed above. Our review of the neoclassical model emphasizes that it is in fact not a model of ongoing growth, since it implies that per capita output rates will approach constant values in the absence of exogenous (therefore unexplained) technological progress. Several analytical results are exposited, including the distinction between golden rule and optimal steady states. Following this review, it is argued that the neoclassical approach not only fails to provide an explanation of everlasting steady-state growth, but also cannot plausibly explain actual observed cross-country growth rate differences by reference to transitional (i.e., non-steady-state) episodes. It can, with the inclusion of human capital inputs, explain a substantial portion of observed cross-country differences in income levels, but there are some questionable aspects of this accomplishment and, in any event, explaining levels is not the main task of a theory of growth.

The endogenous growth literature attempts to provide explanations for ongoing, steady-state growth in per capita output values and consequently for growth rate differences across countries. Three types of endogenous growth models are presented, featuring (i) externalities resulting from linked capital-and-knowledge accumulation, (ii) accumulation of human capital (i.e., individuals' workplace skills), and (iii) continuing growth in the stock of existing productive "designs," with the entire stock facilitating the creation of additional designs (that are produced in response to private rewards). The last of these types seems most plausible as a mechanism capable of generating long-lasting growth. The likelihood of obtaining steady-state (never-ending but non-explosive) growth from any of the models seems very small, however, since such a result would require highly special (zero measure) parameter values. The endogenous growth approach seems fruitful, nevertheless, as it can in principle rationalize long-lasting growth and growth rate differences across economies and will indicate with reasonable accuracy the effects of changes in policy, tastes, or technology that alter the steady-state capital/labor ratio.

APPENDIX A

The purpose here is not to furnish rigorous mathematical proofs but instead to provide some intuition concerning the role and nature of transversality conditions in infinite horizon optimization problems. Let us proceed in the context of the problem of Section 1, to maximize (1) subject to constraint (3). We begin with a *T*-period *finite* horizon version for which the Lagrangian expression is

$$\begin{aligned} L_1 = {} & u(c_1) + \beta u(c_2) + \ \ldots \ + \beta^{T-1}u(c_T) + \lambda_1[f(k_1) - c_1 - (1+\nu)k_2 \\ & + (1-\delta)k_1] + \beta\lambda_2[f(k_2) - c_2 - (1+\nu)k_3 + (1-\delta)k_2] \\ & + \ \ldots \ + \beta^{T-1}\lambda_T[f(k_T) - c_T - (1+\nu)k_{T+1} + (1+\delta)k_T]. \end{aligned} \tag{A1}$$

For $t = 1, 2, \ \ldots, \ T$ we have the first-order conditions

$$u'(c_t) - \lambda_t = 0 \tag{A2}$$

$$-(1+\nu)\lambda_t + \beta\lambda_{t+1}[f'(k_{t+1}) + 1 - \delta] = 0. \tag{A3}$$

In addition there is the derivative with respect to k_{T+1}, $\partial L_1/\partial k_{T+1} = -\lambda_T \beta^{T-1}$ $(1+\nu)$. If the problem were such that one could be assured of a positive solution for k_{T+1}, as is the case for $c_1, \ldots, c_T$ and $k_2, \ldots, k_T$, then one might be inclined to set this partial equal to zero. But of course the household would like for k_{T+1} to be negative and very large, since that would permit c_T to be very large. Thus the inherent constraint $k_{T+1} \geq 0$ becomes relevant and leads to the two-part Kuhn-Tucker condition

$$-\lambda_T \beta^{T-1}(1+\nu) \leq 0 \qquad - k_{T+1}[\lambda_T \beta^{T-1}(1+\nu)] = 0. \tag{A4}$$

Since $\lambda_1, \ldots, \lambda_T$ will by (A2) be strictly positive, the first of these is irrelevant, and the second implies that $k_{T+1} = 0$.

Now consider the infinite horizon version of the same problem by letting $T \rightarrow \infty$. Heuristically we again have conditions (A2) and (A3), relevant for all $t = 1, 2, \ldots$. And in place of (A4) we now have the *TC*

$$\lim_{T\rightarrow\infty} k_{T+1}\beta^{T-1}\lambda_T = 0. \tag{A5}$$

Here the interpretation is that the present value of k_{T+1} in marginal utility units must approach zero as T grows without bound. Since $\beta^{T-1}\rightarrow 0$, this does not require that $k_{T+1}\rightarrow 0$.

Note that it is fortunate that the *TC* is available, for without it (or some replacement) the two difference equations (A2) and (A3) could not provide a well-defined path in the infinite horizon case for there is only one relevant initial condition present (i.e., the given value of k_1). This, then, is the role of the *TC* condition, to provide an additional side condition for starting up the solution sequence $c_1, c_2, \ldots, k_2, k_3, \ldots$. It serves to prevent the optimizing agent from starting on paths that satisfy (A2) and (A3) but lead to negative values of k_t or to wastefully large accumulations of assets that are *never* turned into consumption.

For the infinite horizon problem at hand, it is the case that (subject to the Kuhn-Tucker "constraint qualification") conditions (A2) and (A3) for $t = 1, 2, \ldots$ and condition (A5) are necessary and jointly sufficient for optimality. There are a few exceptional setups with concave objective functions and convex constraint sets, and lots of differentiability, for which the *TC* is not necessary for optimality. But in most infinite horizon problems, the *TC* is also necessary—as is shown by Weitzman (1973).

APPENDIX B

Here we show that if the production function in per capita terms is

$$y = f(n, k, t) \tag{B1}$$

and is homogeneous of degree one (HD1) in n and k, then steady-state growth is possible only when the technical progress term involving t is labor augmenting. To begin, let us assume that labor supply is inelastic so that $n = 1$. Then HD1 implies that $\lambda f(1, k, t) = f(\lambda, \lambda k, t)$ for any $\lambda > 0$. So if we define $x = k/y$, then $1 = f(1/y, x, t)$, which permits us to define the function ϕ such that $y = \phi(x, t)$. Calculating the partial derivative with respect to k then gives

$$\partial y/\partial k = \phi_1(x, t)[-ky^{-2}\partial y/\partial k + y^{-1}], \tag{B2}$$

which can be rearranged to yield

$$\partial y/\partial k = \frac{\phi_1(x, t)}{\phi(x, t) + x\phi_1(x, t)}. \tag{B3}$$

Thus for $\partial y/\partial k$ to be independent of t, we must be able to write the right-hand side of (B3) as $c(x)$, say, implying that $\phi_1(x, t) = c(x)[\phi(x, t) + x\phi_1(x, t)]$ or

$$\frac{\phi_1(x, t)}{\phi(x, t)} = \frac{c(x)}{1 - xc(x)} \tag{B4}$$

so that $\phi(x, t)/\phi_1(x, t)$ is independent of t. But that implies that $\phi(x, t)$ can be written as

$$\phi(x, t) = A(t)\psi(x), \tag{B5}$$

say, so $y = A(t)\psi(x)$ and $x = \psi^{-1}[y/A(t)]$. Then $k = xy = y\psi^{-1}[y/A(t)]$ and

$$k/A(t) = [y/A(t)]\psi^{-1}[y/A(t)] \equiv G[y/A(t)]. \tag{B6}$$

Finally, inversion of G yields

$$y/A(t) = G^{-1}[k/A(t)] = g[k/A(t)]. \tag{B7}$$

Thus it must be that $y = f(1, k, t)$ is of the form $y = \tilde{f}(A(t), k)$.

This proof has been adapted from Uzawa (1961). The statement involving (B5) is treated by Uzawa as obvious. Uzawa's proof pertains, it should be noted, to a proposition that is more general than the one proved by Barro and Sala-i-Martin (1995, pp. 54–55) or Solow (1970, pp. 35–37).

REFERENCES

Aghion, Philippe, and Peter Howitt. “A Model of Growth through Creative Destruction,” *Econometrica,* vol. 60 (March 1992), pp. 323–51.

Barro, Robert J., and Xavier Sala-i-Martin. *Economic Growth*. New York: McGraw-Hill, 1995.

__________. “Convergence,” *Journal of Political Economy,* vol. 100 (April 1992), pp. 223–51.

Blanchard, Olivier J., and Stanley Fischer. *Lectures on Macroeconomics*. Cambridge, Mass.: MIT Press, 1989.

Brock, William A. “Comments,” in *Frontiers of Quantitative Economics,* Vol. II. Amsterdam: North Holland, 1974(a).

__________. “Money and Growth: The Case of Long-Run Perfect Foresight,” *International Economic Review,* vol. 15 (1974b), pp. 750–77.

Brock, William A., and Leonard J. Mirman. “Optimal Economic Growth and Uncertainty: The Discounted Case,” *Journal of Economic Theory,* vol. 4 (June 1972), pp. 479–513.

Cass, David. “Optimum Growth in an Aggregative Model of Capital Accumulation,” *Review of Economic Studies,* vol. 32 (1965), pp. 233–40.

Cho, Dongchul, and Stephen Graham. “The Other Side of Conditional Convergence,” *Economics Letters,* vol. 50 (February 1996), pp. 285–90.

Denison, Edward F. *Why Growth Rates Differ*. Washington, D.C.: The Brookings Institution, 1967.

Domar, Evsey. “Expansion and Employment,” *American Economic Review,* vol. 37 (1947), pp. 343–55.

Goodfriend, Marvin, and John McDermott. “Early Development,” *American Economic Review,* vol. 85 (March 1995), pp. 116–33.

Grossman, Gene M., and Elhanan Helpman. “Endogenous Innovation in the Theory of Growth,” *Journal of Economic Perspectives,* vol. 8 (Winter 1994), pp. 23–44.

__________. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press, 1991.

Hahn, Frank H. “Neoclassical Growth Theory,” in John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, Vol. 3. New York: Stockton Press, 1987.

Harrod, R. F. “An Essay in Dynamic Theory,” *Economic Journal,* vol. 49 (1939), pp. 14–33.

Islam, Nazrul. “Growth Empirics: A Panel Data Approach,” *Quarterly Journal of Economics,* vol. 110 (November 1995), pp. 1127–70.

Jones, Charles I. "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics,* vol. 110 (May 1995), pp. 495–525.

King, Robert G., and Ross Levine. "Finance, Entrepreneurship, and Growth," *Journal of Monetary Economics,* vol. 32 (December 1993), pp. 513–42.

King, Robert G., and Sergio Rebelo. "Transitional Dynamics and Economic Growth in the Neoclassical Model," *American Economic Review,* vol. 83 (September 1993), pp. 908–31.

__________. "Public Policy and Economic Growth: Developing Neoclassical Implications," *Journal of Political Economy,* vol. 98 (October 1990), pp. S126–S150.

Koopmans, Tjalling C. "On the Concept of Optimal Economic Growth," in *Study Week on the Econometric Approach to Development Planning*. Amsterdam: North-Holland Pub. Co., 1965.

Levine, Ross, and David Renelt. "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review,* vol. 82 (September 1992), pp. 942–63.

Lucas, Robert E., Jr. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review Papers and Proceedings,* vol. 80 (May 1990), pp. 92–96.

__________. "On the Mechanics of Economic Development," *Journal of Monetary Economics,* vol. 22 (July 1988), pp. 3–42.

Mankiw, N. Gregory. "The Growth of Nations," *Brookings Papers on Economic Activity,* 1:1995, pp. 275–310 and 324–26.

Mankiw, N. Gregory, David Romer, and David N. Weil. "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics,* vol. 107 (May 1992), pp. 407–37.

McCallum, Bennett T. "Macroeconomics after Two Decades of Rational Expectations," *Journal of Economic Education,* vol. 25 (Summer 1994), pp. 219–34.

__________. "Inflation: Theory and Evidence," in Benjamin M. Friedman and Frank H. Hahn, eds., *Handbook of Monetary Economics,* Vol. II. Amsterdam: North-Holland Pub. Co., 1990.

__________. "Real Business Cycle Models," in Robert J. Barro, ed., *Modern Business Cycle Theory*. Cambridge, Mass.: Harvard University Press, 1989.

Meade, James E. *A Neoclassical Theory of Economic Growth*. London: Allen & Unwin, 1962.

Mulligan, Casey B., and Xavier Sala-i-Martin. "Transitional Dynamics in Two-Sector Models of Endogenous Growth," *Quarterly Journal of Economics,* vol. 108 (August 1993), pp. 739–73.

Phelps, Edmund S. "The Golden Rule of Accumulation: A Fable for Growthmen," *American Economic Review,* vol. 51 (1961), pp. 638–43.

Ramsey, Frank P. "A Mathematical Theory of Saving," *Economic Journal,* vol. 38 (1928), pp. 543–49.

Rebelo, Sergio. "Long-Run Policy Analysis and Long-Run Growth," *Journal of Political Economy,* vol. 99 (June 1991), pp. 500–21.

Rivera-Batiz, Luis A., and Paul M. Romer. "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics,* vol. 106 (May 1991), pp. 531–55.

Romer, Paul M. "The Origins of Endogenous Growth," *Journal of Economic Perspectives,* vol. 8 (Winter 1994), pp. 3–22.

__________. "Endogenous Technological Change," *Journal of Political Economy,* vol. 98 (October 1990), pp. S71–S102.

__________. "Capital Accumulation in the Theory of Long-Run Growth," in Robert J. Barro, ed., *Modern Business Cycle Theory*. Cambridge, Mass.: Harvard University Press, 1989.

__________. "Crazy Explanations for the Productivity Slowdown," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1987*. Cambridge, Mass.: MIT Press, 1987.

__________. "Increasing Returns and Long-Run Growth," *Journal of Political Economy,* vol. 94 (October 1986), pp. 1002–37.

Solow, Robert M. "Perspectives on Growth Theory," *Journal of Economic Perspectives,* vol. 8 (Winter 1994), pp. 45–54.

__________. *Growth Theory: An Exposition*. Oxford: Clarendon Press, 1970.

__________. "Comment on the Golden Rule," *Review of Economic Studies,* vol. 29 (1962), pp. 255–57.

__________. "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics,* vol. 70 (1956), pp. 65–94.

Summers, Robert, and Alan Heston. "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950–1985," *Review of Income and Wealth,* vol. 34 (March 1988), pp. 1–25.

Swan, Trevor W. "Economic Growth and Capital Accumulation," *Economic Record,* vol. 32 (1956), pp. 334–61.

Uzawa, H. "Optimal Technical Change in an Aggregative Model of Economic Growth," *International Economic Review,* vol. 6 (1965), pp. 18–31.

__________. "Neutral Inventions and the Stability of Growth Equilibria," *Review of Economic Studies,* vol. 28 (1961), pp. 117–24.

Weitzman, Martin L. "Duality Theory for Infinite Horizon Convex Models," *Management Science,* vol. 19 (March 1973), pp. 783–89.

Some Not-So-Unpleasant Monetarist Arithmetic

Michael Dotsey

The motivation for this article is the well-known and controversial work of Sargent and Wallace (1981), who show the potential importance of the government's budget constraint for the behavior of nominal variables. The government's lifetime budget constraint can place restrictions on the behavior of future money growth and thus influence current economic magnitudes through expectational channels. Some of Sargent and Wallace's results are indeed striking, and indicate that tight monetary policy can lead to the unpleasant outcome of both higher expected inflation and a higher price level. Theoretically, they highlight important intertemporal considerations, but one wonders if these considerations are quantitatively meaningful in reality.

In order to investigate the quantitative significance of monetarist arithmetic, I develop a dynamic stochastic model in which the government's budget constraint has nontrivial implications. In particular, I use the methodology of Dotsey (1994) and Dotsey and Mao (1996). Here both money growth and taxes are stochastic, but one or both must endogenously respond to government debt if the government is to maintain budget balance. When the monetary authority does not respond to debt, I term its policy *independent* monetary policy, and when it does respond to debt, I call it *dependent* monetary policy. The primary focus of the article is on the behavior of nominal variables and involves a comparison of their behavior when the monetary authority is and is not independent. The main result is that, for reasonable parameterizations, and when the tax authority responds to the level of debt, the underlying nominal behavior of the economy does not significantly depend upon whether the monetary authority reacts to government financing considerations. In this sense the monetarist arithmetic is not so unpleasant.

I would like to thank Zvi Hercowitz, Robert Hetzel, Tom Humphrey, Peter Ireland, Alan Stockman, and Alex Wolman for a number of useful suggestions and Jed DeVaro for research assistance. The views expressed in this article are those of the author and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

The considerations addressed in this article are also similar to those discussed in Aiyagari and Gertler (1985), Leeper (1991), and Woodford (1995). These authors are all concerned with the relationship between the behavior of money and nominal variables in models that emphasize the importance of the government's budget constraint. The article proceeds as follows. Section 1 reviews the key elements of the above-cited papers that are most relevant to the experiments carried out in this article. Section 2 describes the behavior of taxes and money growth as well as the underlying economic model. Section 3 investigates parameterizations that produce the spectacular example presented in Sargent and Wallace, in which current inflation increases under contractionary monetary policy. These parameterizations do not accurately characterize the U.S. economy and the spectacular example does not occur under realistic behavioral assumptions. Section 4 analyzes a model economy that is more consistent with that of the United States. For this case, data are generated under both independent and dependent monetary policy rules. Only minor differences are found in the behavior of nominal variables. Section 5 concludes.

1. LITERATURE SURVEY

Sargent and Wallace (1981) present a model economy that satisfies the monetarist assumptions that the monetary base is closely connected to the price level and that the monetary authority can raise seignorage through money creation. They show that under certain conditions the monetary authority's ability to control inflation is limited. The major condition responsible for this result is the exogeneity of the process for the government's deficit. Given this exogeneity, tight money today leads to higher inflation in the future and may even lead to higher inflation today.

Leeper (1991) extends the Sargent and Wallace analysis to a stochastic environment. Leeper's model is, however, somewhat different in that the monetary authority uses an interest rate instrument and reacts, in some cases, to the rate of inflation. The monetary authority does not react specifically to debt as it does in the model presented below. Also, another distinction between Leeper's model and my model is that I use money as the policy instrument.[1] This modeling is more directly related to Sargent and Wallace.

Like Leeper, I find that when monetary policy is independent, fiscal disturbances in the form of lump sum taxes have no influence on either nominal or real variables. Also, as in Leeper, when the monetary authority responds to debt, fiscal disturbances do affect the economy. Both the price level and the nominal

[1] One could always make the money supply rule more realistic by including elements of interest rate smoothing. For a detailed discussion on interest rate instruments in a rational expectations model, see Boyd and Dotsey (1996).

interest rate are positively related to debt. Contrary to his analysis, however, I find that a monetary contraction does not generally cause the nominal interest rate to rise. This difference is a result of the more general stochastic process that governs monetary policy, a process that allows for varying degrees of persistence in policy. A current decline in money growth need not be offset by an immediate increase in next period's money growth as occurs in one of the examples emphasized in his paper.

The analysis here is also related to the work of Aiyagari and Gertler (1985) and Woodford (1995). These authors indicate that when monetary policy is influenced by the government's intertemporal budget constraint, nominal variables are sensitive to debt. In contrast to the results in Aiyagari and Gertler, the model considered here indicates that the nominal interest rate and inflation are not necessarily higher when monetary policy depends on the level of government debt. The difference in results occurs because the stochastic process for money growth implies that low money growth rates are more likely when the debt is low. Thus, the possible values of the nominal interest under a dependent monetary policy span the values that occur when monetary policy is independent.

Much of the work in this area adopts the confusing terminology of Ricardian and non-Ricardian monetary policy. I choose not to use this terminology and instead use the terms independent and dependent monetary policy. As Aiyagari and Gertler point out, Ricardian monetary policies do not necessarily imply that the Ricardian equivalence theorem holds. Also, different authors appear to have somewhat different interpretations of what constitutes a Ricardian monetary regime. Aiyagari and Gertler adopt Sargent's (1982) definition, which refers to whether or not government bonds are fully backed by taxes. That is, the fiscal authority commits to a tax process with a present discounted value equal to the present discounted value of government spending and the current value of outstanding government debt. Woodford's definition is somewhat different. It refers to the transversality condition on debt and whether that condition holds independently. Thus, in Woodford's terminology, a model can be Ricardian even though seignorage revenues respond to changes in government debt. His definition, therefore, allows money to respond to debt in a Ricardian world. In all the models considered below, the real value of the debt is bounded, and hence its present discounted value approaches zero. Yet debt does have real and nominal effects in some of the models. Therefore, I prefer to use the terminology of independent and dependent monetary policy.

2. THE MODEL

In this section I depict the fiscal and monetary policy process as well as the technology and behavioral assumptions that characterize the economic environment relevant to this study.

Fiscal and Monetary Policy Processes

The monetary and fiscal policy processes analyzed here are stochastic and satisfy the government's budget constraint. Monetary policy is defined over changes in base growth rather than in terms of setting the nominal interest rates. This depiction of policy is consistent with previous literature relating deficit finance and inflation.[2] This modeling of monetary policy also allows one to investigate the consequences of money growth's dependence on debt while incorporating the empirically relevant behavior of taxes responding to debt (see Bohn [1991]). Interest rate smoothing could easily be incorporated by allowing the monetary authority to respond to unexpected movements in the nominal interest rate without affecting the existence or uniqueness of the solutions (see Boyd and Dotsey [1996]).

Money, M_t, is introduced through open market operations and behaves according to

$$M_t = M_{t-1}(1 + \eta_t), \tag{1}$$

where η_t is the stochastic rate of money growth.

Taxes are proportional to output. Thus, tax revenues, T_t, are given by

$$T_t = \tau_t Y_t, \tag{2}$$

where Y_t is current nominal output and τ_t is the current tax rate. The government's nominal debt, B_{t+1}, therefore, follows

$$P_t^B B_{t+1} = G + B_t - T_t - \eta_t M_{t-1}, \tag{3}$$

where G represents fixed transfer payments, and P_t^B is the price of a bond at date t that pays one dollar at date $t+1$. The real value of debt relative to ouput, $B_{t+1}/Y_{t+1} = b_{t+1}$, can be written as

$$b_{t+1} = R_t[g + b_t - \tau_t - (\eta_t/(1 + \eta_t))(M_t/Y_t)](Y_t/Y_{t+1}), \tag{4}$$

where R_t is the gross one period nominal interest rate.

A sufficient condition for the government to obey its lifetime budget constraint is for the tax and money growth processes to behave so that the debt-to-GNP ratio is bounded. For that to happen, either one or both processes must depend on government debt. When money and tax rates respond to debt they are modeled as two-state Markov processes with endogenous transition probabilities. Thus the processes are time varying. In particular, let the transition probabilities for taxes be

$$prob(\tau_{t+1} = \tau_l | \tau_t = \tau_l) = \begin{array}{ll} 1 & if\ b_t < 0 \\ (1 - \phi b_t)^{1/\nu} & if\ 0 \leq b_t \leq 1/\phi \\ 0 & if\ b_t > 1/\phi \end{array} \tag{5}$$

[2] For example see Sargent and Wallace (1981), Liviatan (1984), McCallum (1984), Drazen (1985), Aiyagari and Gertler (1985), and Woodford (1995).

$$prob(\tau_{t+1} = \tau_h | \tau_t = \tau_h) = \begin{cases} 0 & \textit{if } b_t < 0 \\ \phi b_t^{1/\nu} & \textit{if } 0 \leq b_t \leq 1/\phi, \\ 1 & \textit{if } b_t > 1/\phi \end{cases}$$

and those for money growth be

$$prob(\eta_{t+1} = \eta_l | \eta_t = \eta_l) = \begin{cases} 1 & \textit{if } b_t < 0 \\ (1 - \phi b_t)^{1/\psi} & \textit{if } 0 \leq b_t \leq 1/\phi \\ 0 & \textit{if } b_t > 1/\phi \end{cases} \tag{6}$$

$$prob(\eta_{t+1} = \eta_h | \eta_t = \eta_h) = \begin{cases} 1 & \textit{if } b_t < 0 \\ \phi b_t^{1/\psi} & \textit{if } 0 \leq b_t \leq 1/\phi, \\ 0 & \textit{if } b_t > 1/\phi \end{cases}$$

where the subscripts l and h refer to low and high, respectively.

As long as the debt-to-GNP ratio rises when both taxes and money growth rates are low and falls when tax rates and money growth rates are high, the debt-to-GNP ratio is bounded and will only rarely lie outside the interval $[0,1/\phi]$. As b_t approaches $1/\phi$, both taxes and money growth will be high with probability one. Similarly, as b_t approaches 0, both taxes and money growth will be low. The parameters ν and ψ control the persistence of the two processes. For any given debt-to-GNP ratio, as ν and ψ increase, the probability of remaining in the current state increases. The above specifications of policy imply that current realizations of taxes and money growth have implications for the entire future path of policy through their effect on debt.

It is important to note that both the unconditional mean and the persistence of money growth do not depend on the debt-to-GNP ratio. If instead the monetary authority had no control over the mean and over persistence of money growth rates, it would be trivial to show that the nominal behavior of the economy depended on fiscal policy. Instead I explore the more difficult question of whether conditional dependence of monetary policy on debt affects nominal magnitudes.

Alternatively, cases in which one of the processes is invariant to debt can be analyzed. I allow for the invariance of tax rates when investigating parameterizations that yield the spectacular case of Sargent and Wallace. I also compare the nominal behavior of an economy when the money growth process does and does not depend on debt.

Technology and Preferences

Since I am primarily concerned with the nominal behavior of the economy, I treat real output as fixed at a constant level, y. Agents derive utility from consumption, c, and real balances, m. Specifically, they maximize

$$U = \max E \sum_{t=0}^{\infty} \beta^t [c_t^{-\rho} + \theta m_t^{-\rho}]^{-1/\rho}$$

subject to the per-period budget constraint

$$M_t^d + P_t^B B_{t+1}^d + p_t c_t \leq (1 - \tau_t) p_t y_t + M_{t-1} + B_t + G_t,$$

where p is the price level and E is the expectations operator conditional on time 0 information. This specification of preferences allows one to look at a range of interest elasticities of money demand and to freely parameterize the ratio m/c.

The first-order conditions determining the demand for money and the optimal consumption-saving decision by individuals are given by

$$\left(\frac{m_t}{c_t}\right) = \left(\frac{\theta(1 + r_t)}{r_t}\right)^{1/(1+\rho)} \tag{7}$$

and

$$\left(\frac{1}{1 + r_t}\right) \left[1 + \theta \left(\frac{c_t}{m_t}\right)^{\rho}\right]^{-(1/\rho)-1} =$$

$$\beta E_t \left\{ \left[1 + \theta \left(\frac{c_{t+1}}{m_{t+1}}\right)^{\rho}\right]^{-(1/\rho)-1} \left(\frac{p_t}{p_{t+1}}\right) \right\}, \tag{8}$$

where r_t is the net nominal interest rate and E_t is the expectations operator conditional on time t information. Agents are assumed to know all contemporaneously dated variables as well as the nominal value of end-of-period government debt. Equation (7) implies that money demand is unit elastic with respect to the scale variable consumption and has an interest elasticity of $-(1/(1 + \rho))(1/(1 + r))$. Also, for a given steady-state interest elasticity the parameter θ determines the velocity of money. For example, as ρ goes to infinity $m = c$ and the model approaches a cash-in-advance specification, while as ρ goes to zero money demand becomes unit elastic. Equation (8) governs the optimal consumption-saving decision. Note that real money balances influence intertemporal consumption choices through their effect on the marginal utility of consumption. Higher current real money balances increase the marginal utility of consumption and, holding expected future consumption and money balances constant, imply a higher real interest rate.

Equilibrium

The equilibrium conditions for this economy are

$$c_t = y, \tag{9}$$

$$M_t/p_t = m_t, \tag{10}$$

and

$$B^d_{t+1} = B_{t+1}, \tag{11}$$

along with the first-order conditions (7) and (8), the agent's budget constraint, and the government's budget constraint (3). The three equations (9) – (11) merely state that demand equals supply in the goods, money, and bond markets, respectively. Substituting (7) into (8) and using (9) and (10), these four equations can be used to derive a functional equation for the nominal interest rate. Combined with the law of motion for real government debt (equation [4]) and the stochastic processes for taxes and money growth, the equilibrium functions for the nominal interest rate $r(b, \tau, \eta)$ and next period's real debt $b'(b, \tau, \eta, \tau', \eta')$ can be solved (where the symbol " $'$ " refers to next period value of a variable). Using the solution for r, one can readily derive the solutions for real balances, prices, and expected inflation. Thus, equilibrium is a set of functions for r, b', p, c, and m that satisfy equations (4), and (7) – (11). Because there is no closed form solution, I obtain the solution numerically by solving for the relevant functions along a grid of real debt levels and at the values of taxes and money growth.

3. THE SPECTACULAR CASE

In their spectacular example, Sargent and Wallace showed that it is possible for lower current money growth to cause both higher expected inflation and a higher price level. The higher expected inflation occurs because lower money growth increases the government's indebtedness, implying that money growth must be, on average, higher in the future. The higher expected future money growth leads to higher expected inflation. Also, the higher expected inflation reduces the current demand for real balances. If the reduction in demand is large enough, the price level must rise to clear the money market. As Drazen (1985) points out, this latter result requires an interest elasticity of at least one in absolute value. Using the model of the previous section, we are able to produce a spectacular example. To do so requires some additional assumptions that are extreme when compared to actual economic behavior in the United States. Besides the unusual responsiveness of money demand to nominal interest rates, the ratio m/c must be higher than one observes in the data, and rates of money growth in the high money growth state must be somewhat higher than commonly observed in the post-war United States. Both of these counterfactual parameterizations must be made in order to bound the debt-to-GNP ratio. That is, both the tax base and the tax rate on money balances in the high money growth state must be sufficiently high to pay off the increased debt burden that accrues in the low money growth state.

The policy functions for the spectacular example are depicted in Figure 1. To produce this example, the interest elasticity of money demand was set at -1, the steady-state ratio m/c was set at 0.25, and the low and high money growth rates were selected to be -1.25 percent and 13.5 percent, respectively. Also, money growth is quite persistent with $\psi = 5$. This parameterization corresponds to an autocorrelation of roughly 0.65. The model is calibrated at an annual frequency and $\beta = 0.98$. With these parameter values, the debt to GNP ratio lies between -0.46 and 0.58. As shown in the top panel of Figure 1, the nominal interest rate is higher when money growth is low, even though the money growth process is characterized by substantial persistence.[3] This persistence is depicted in the middle panel of Figure 1, which shows that agents expect substantially higher future money growth when money growth is currently high. One also observes, in the bottom panel, that the price level is higher when money growth is low.

To highlight the necessity of the high interest elasticity, I also examine the case in which the interest elasticity is -0.20. This value is representative of money demand in the United States (see Dotsey [1988]). As shown in Figure 2, both the nominal interest rate and the price level are higher when current money growth is high. Thus, the spectacular case is reversed. Additional insight into the important role that the interest elasticity plays in producing the spectacular result can be found in Table 1. This table shows the interest rates that would occur under the high and low money growth rates for various elasticities of money demand. These interest rates are obtained from the following thought experiment. Suppose that the transition probability for money growth is 0.8 and independent of the debt. Also suppose that the tax authority adjusts taxes to bound the level of debt. Then high current money growth always implies higher future money growth. The interest rates that solve equation (8) can be calculated for various values of the interest elasticity and a value of $m/c = 0.25$. As depicted in the table, the difference between the nominal interest rate when money growth is high as opposed to when it is low narrows substantially as the interest elasticity increases in absolute value.

The narrowing of the spread between interest rates under high and low money growth along with the implication that current low money growth leads to higher average future money growth produce the spectacular example. Furthermore, the high interest elasticity implies that money demand falls by more than one-for-one with the money supply and the price level must rise to equilibrate the money market.

[3] The policy function for expected inflation is qualitatively identical to that of the nominal interest rate. It is, therefore, omitted.

Figure 1

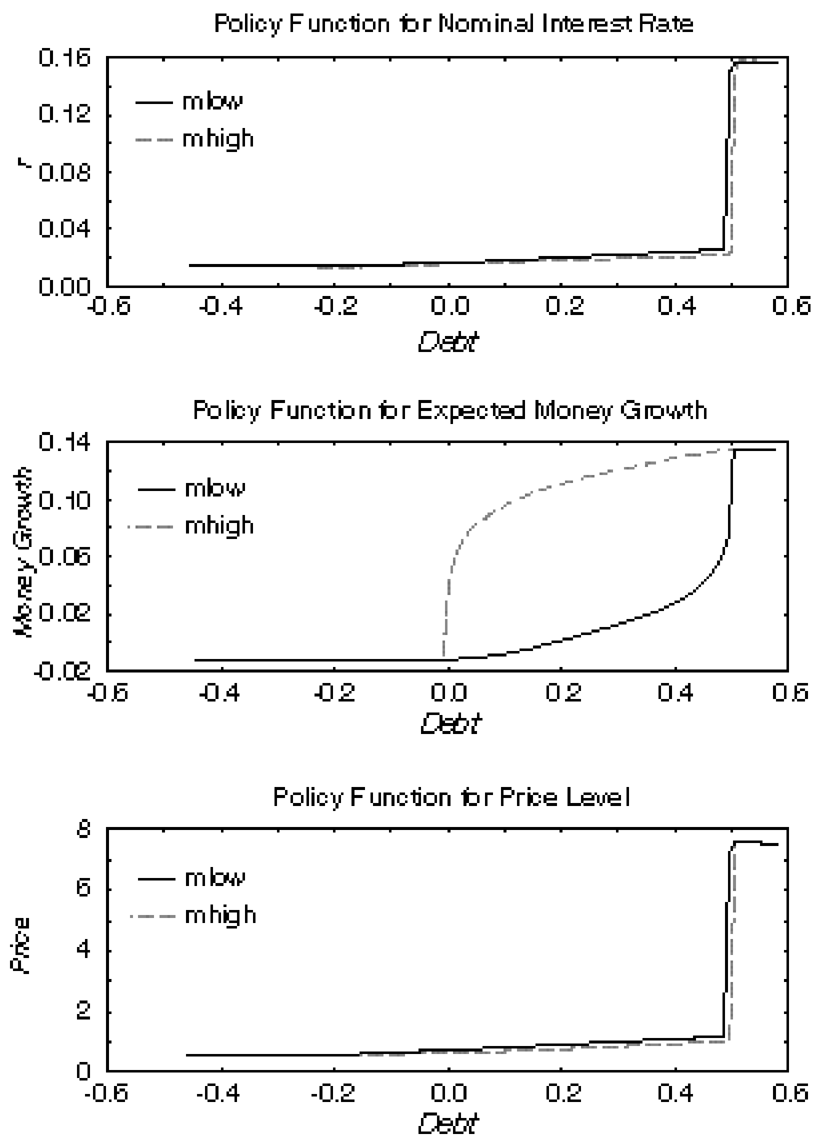

Policy Function for Nominal Interest Rate
r
0.16
0.12
0.08
0.04
0.00
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Debt
mlow
mhigh
Policy Function for Expected Money Growth
Money Growth
0.14
0.10
0.06
0.02
-0.02
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Debt
mlow
mhigh
Policy Function for Price Level
Price
8
6
4
2
0
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Debt
mlow
mhigh

Figure 2

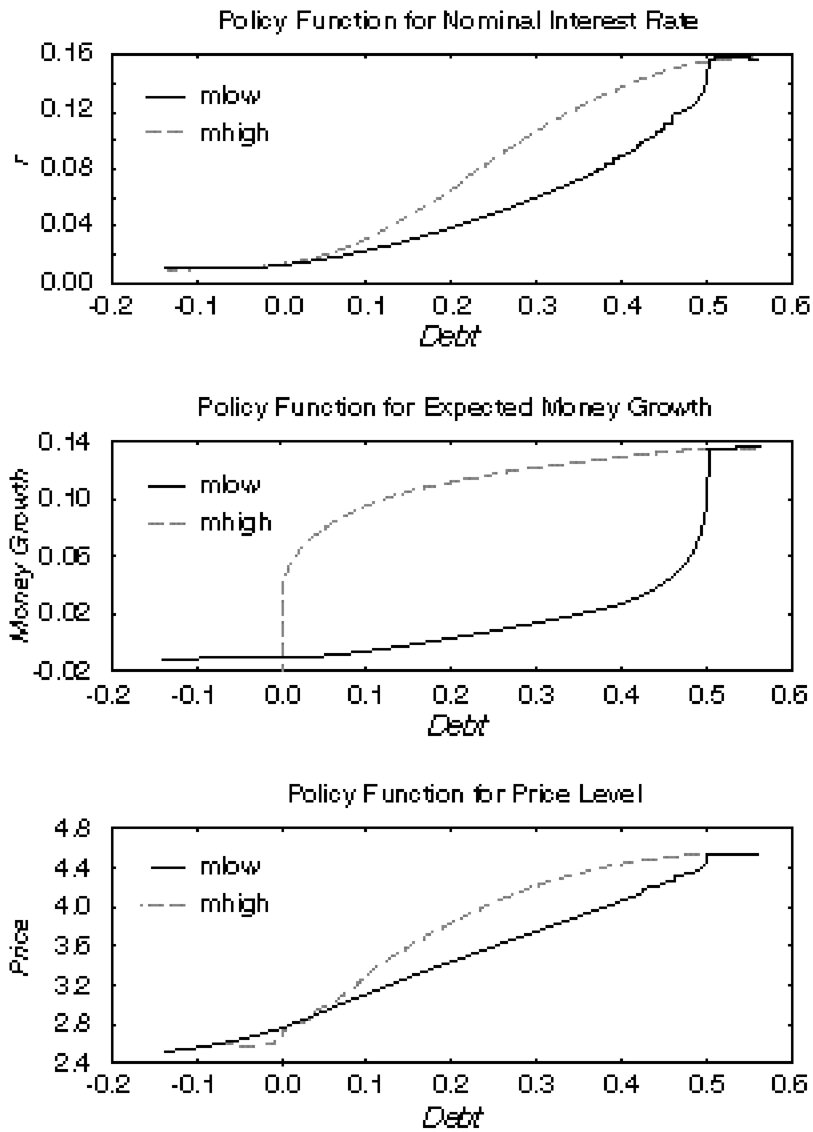

Policy Function for Nominal Interest Rate
mlow
mhigh
r
0.16
0.12
0.08
0.04
0.00
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Debt
Policy Function for Expected Money Growth
mlow
mhigh
Money Growth
0.14
0.10
0.06
0.02
-0.02
Debt
Policy Function for Price Level
mlow
mhigh
Price
4.8
4.4
4.0
3.6
3.2
2.8
2.4
Debt

Table 1 Interest Elasticity Effects on Interest Rates

Interest Elasticity of Money Demand	R_l	R_h
−0.10	0.051	0.119
−0.25	0.063	0.100
−0.50	0.068	0.092
−0.75	0.071	0.088
−1.00	0.072	0.086

Note: In this example money growth takes on the values −0.0055 and 0.142, and the transition probability of remaining in the same state is 0.8.

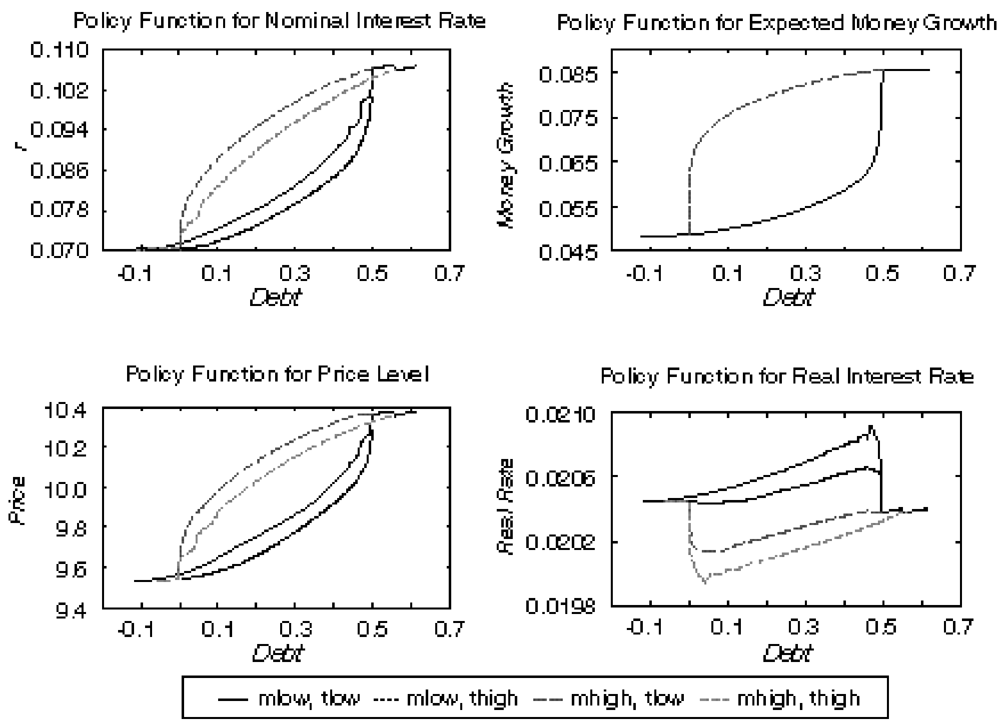
4. MONETARIST ARITHMETIC IN A CALIBRATED MODEL

To explore the empirical significance of monetarist arithmetic, I calibrate two related models and compare the behavior of nominal variables and expected lifetime utility. In both models, endowment is fixed at one unit and government transfers are equal to 0.189. The models are calibrated to produce a steady-state interest elasticity of money demand of −0.20 and a ratio of money to consumption of 0.10. At an annual frequency, both of these values are consistent with U.S. data. The discount factor is set at 0.98, implying a steady-state real interest rate of two percent. The money growth rates are 0.049 and 0.086, while the tax rates are 0.17 and 0.229. The debt-to-GNP ratio lies within the interval [−0.12, 0.62]. The two models are similar in that the tax authority responds to the debt-to-GNP ratio as described in equation (5). The persistence parameter, υ, is set at 3.5. The first model considered depicts a dependent monetary policy as described by equation (6) with a persistence parameter, ψ, equal to 5.0. The second model allows monetary policy to be independent and the transition probability of remaining in the same state of money growth is 0.85. These parameterizations of tax rates and money growth produce realizations that are consistent with U.S. experience over the period 1961 to 1995.

Policy Functions

The policy functions for the case in which monetary policy responds to debt are depicted in Figure 3. One sees that expected money growth is positively related to government debt. As government debt approaches one half of GNP, the probability of both taxes and money growth being high approaches one. Similarly, as debt approaches zero, the probability of both taxes and money growth being low approaches one. Thus, the policy functions for the different states converge as debt approaches its upper and lower bounds. The positive

Figure 3



relationship between money growth and debt leads to a positive relationship between debt and both the nominal interest rate and the price level. The persistence of the money growth process also implies that higher rates of money growth lead to higher nominal interest rates and a higher price level. Also, for any given rate of money growth, higher taxes imply lower nominal interest rates and lower prices. This result occurs because higher taxes reduce the debt and, therefore, imply a lower average rate of future money growth. Consequently, like Leeper (1991) and Aiyagari and Gertler (1985), the model implies that fiscal policy affects nominal variables through its influence on the path of future money growth.

The real interest rate in this model is also related to the level of debt, monetary policy, and fiscal policy. Higher rates of money growth are associated with lower real balances and with a lower marginal utility of consumption. In the case where money growth is high, the expected value of next period's money growth will be less than its current rate. Hence, next period's real balances are expected to be higher and next period's expected marginal utility will be greater than current marginal utility. This relationship between current and expected future marginal utility implies that the real interest rate will be lower when money growth is currently high. Also, for any given rate of money growth, higher taxes imply lower expected money growth in the future. This lower expected future money growth is associated with higher expected future real balances and a higher expected future marginal utility of consumption. Thus, the real interest rate is lower when taxes are high. Similarly, because lower debt implies lower future money growth, the real interest rate is low when debt is low.

When monetary policy is independent, the policy functions look very different. As shown in Figure 4, debt and fiscal policy no longer have any effects on economic variables. This invariance occurs because taxes and debt do not influence the path of money growth. The policy functions, therefore, take on only two values, one is associated with low money growth, and the other with high money growth. As in the previous case, high money growth results in higher interest rates and a higher price level. Comparison of Figures 3 and 4 reveals that the interest rate and the price level are not necessarily always lower under independent monetary policy. This latter result contrasts with the findings of Aiyagari and Gertler (1985) and occurs because conditional on the debt being low, money growth is much more likely to be low with a dependent monetary policy. Their results, therefore, are sensitive to the actual stochastic specification of policy.

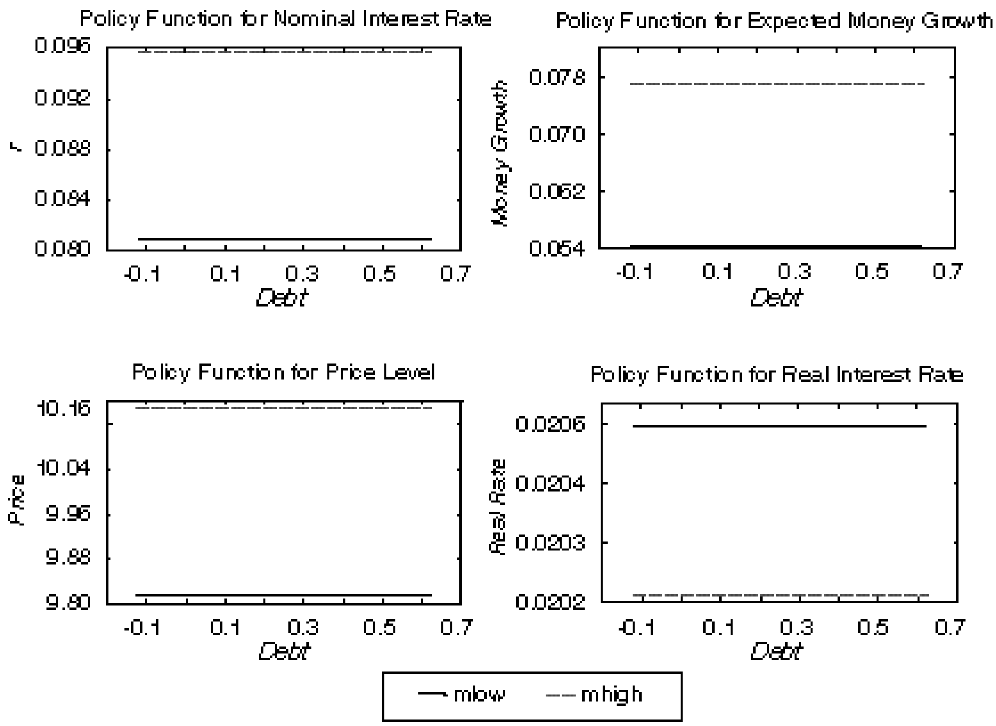
Further Comparison of the Two Policies

To further contrast economic consequences of independent and dependent monetary policies, I examine time series from model simulations. Each simulation is 100 periods long and the results are averaged over 500 simulations. With respect to dependent monetary policy, the simulated tax process has a mean of 0.187 and a standard deviation of 0.027. These values compare with an actual mean of average tax revenues over the period 1961 to 1995 of 0.189 and standard deviation of 0.0074. The simulated tax process is somewhat more variable than the historical series on average tax rates but shows about the same degree of variability as the average marginal tax rate series computed using the methodology of Barro and Sahasakul (1986).[4] The mean of money growth is 0.06 and its standard deviation is 0.017. These values are similar to the actual values for the growth rate of the monetary base, which had a mean of 0.067 and a standard deviation of 0.020.

A further comparison between the generated processes for taxes and money growth and their empirical counterparts can be obtained by examining the

[4] The Barro and Sahasakul series was taken from DRI, and its standard deviation is 0.0260.

Figure 4



following linear regressions. For generated data under a dependent monetary policy, the regressions are given by

$$\underset{}{\tau_t} = \underset{(0.014)}{0.070} + \underset{(0.082)}{0.561}\tau_{t-1} + \underset{(0.023)}{0.073}b_{t-1}, \tag{12}$$

$$\eta_t = \underset{(0.005)}{0.017} + \underset{(0.109)}{0.700}\eta_{t-1} + \underset{(0.014)}{0.007}b_{t-1}, \tag{13}$$

while for an independent monetary policy the corresponding regressions are

$$\tau_t = \underset{(0.014)}{0.064} + \underset{(0.083)}{0.600}\tau_{t-1} + \underset{(0.023)}{0.067}b_{t-1}, \tag{14}$$

$$\eta_t = \underset{(0.007)}{0.024} + \underset{(0.080)}{0.663}\eta_{t-1} - \underset{(0.015)}{0.010}b_{t-1}, \tag{15}$$

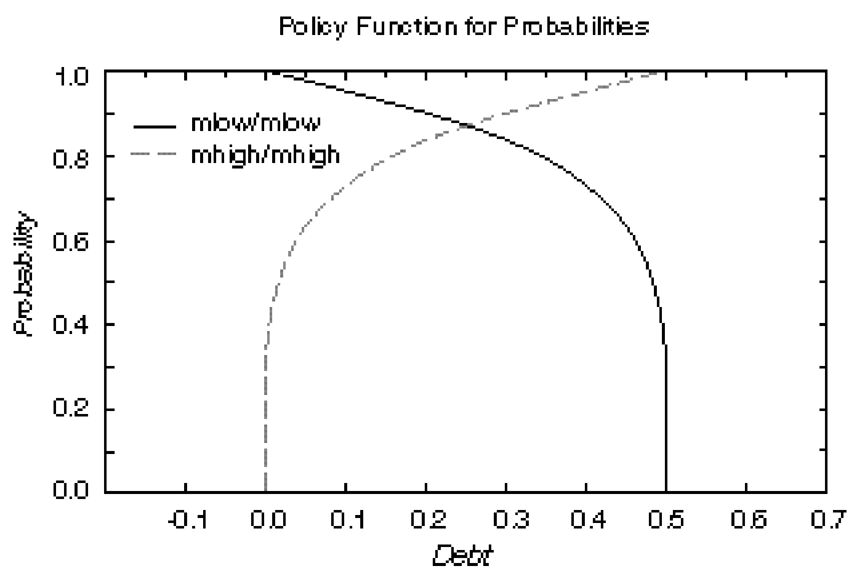
where standard errors are in parentheses. From these regressions alone, it would be difficult to distinguish between the two regimes. All the comparable coefficients differ insignificantly from each other and in neither regime does monetary policy appear to depend on debt. The reason for the insignificant coefficient on debt in the money growth regression (13) is that taxes are doing the bulk of the work in controlling debt. The money growth process is highly persistent (see Figure 5), even at the bounds of the debt space, and taxes are more likely to change states in order to reduce the debt when it is high or increase the debt when it is low. Thus, a simple econometric exercise to test whether seignorage is influenced by debt would fail to uncover this feature in the dependent monetary regime.

Based on actual data over the period 1961 to 1995, the corresponding regressions are

$$\underset{}{\tau_t} = \underset{(0.027)}{0.070} + \underset{(0.14)}{0.62}\tau_{t-1} + \underset{(0.0097)}{0.005}b_{t-1}, \tag{16}$$

$$\eta_t = \underset{(0.015)}{0.028} + \underset{(0.095)}{0.68}\eta_{t-1} - \underset{(0.021)}{0.013}b_{t-1}. \tag{17}$$

Figure 5



With the exception of the insignificance of debt on average tax rates, the regressions on actual and generated data are quite similar.[5] The calibrated processes, therefore, seem to be fairly representative of actual processes.

To compare the influences of monetary policy in the two model economies, I examine impact effects and correlations. For the chosen parameterization the steady-state value of the nominal interest rate is 0.0887, the real interest rate is 0.020, and the price level is $10M$. With low taxes, low money growth, and a dependent monetary policy, the nominal interest rate falls to 0.0807 and the price level declines to 9.810. If taxes were high, then the interest rate would decline to 0.0770 and the price level would fall to 9.719. With an independent monetary policy, low money growth results in a nominal interest rate of 0.0809 and a price level of 9.814. Here, unlike the finding in Aiyagari and Gertler (1985), monetary policy in a dependent regime does affect the price level. Furthermore, although fiscal policy also affects the price level, the effect is much less than one-for-one. For example, under low money growth and high taxes, inflation is 4.3 percent as opposed to 5.0 percent when taxes are low. These contrary results occur for two reasons. One is the explicit stochastic process for policy and the other is the lower interest elasticity of money demand. As the interest elasticity approaches zero, so that $m = c$, prices must move one-for-one with money regardless of the regime.

The preceding example shows that the impact effects of monetary policy differ somewhat across regimes. It remains to ask whether the correlations between policy and nominal variables are very different across regimes. The correlations are depicted in Table 2. The top panel shows the correlation coefficients that occur under a dependent monetary policy and the bottom panel depicts the same correlations under an independent monetary policy. With respect to the correlations between money growth, η, and other nominal variables or between money growth and real balances, the correlations are somewhat smaller under a dependent monetary policy, but by-and-large the correlations are quite similar. The only quantitatively significant differences occur with respect to correlations involving the real interest rate. However, as shown by the policy functions, the real interest rate shows very little variation. Given the presence of measurement error, it would be difficult in practice to identify the monetary regime through the use of data on ex ante real interest rates. Thus, the data generated by the two different models are very much the same.

[5] Although taxes do not seem to respond to debt over the short sample period considered here, fiscal policy does seem to respond to debt over longer time horizons (see Bohn [1991]) and at lower frequencies (see Dotsey and Mao [1996]).

Table 2 Correlation Coefficients

Dependent Monetary Policy

	m	r	π^e	π	rr
η	−0.848	0.854	0.862	0.864	−0.900
m		−0.998	−0.999	−0.759	0.577
r			1.000	0.765	−0.586
π^e				0.770	−0.599
π					−0.742

Independent Monetary Policy

	m	r	π^e	π	rr
η	−1.00	1.00	1.00	0.88	−1.00
m		−1.00	−1.00	−0.88	1.00
r			1.00	0.88	−1.00
π^e				0.88	−1.00
π					0.88

Note: η is money growth, m is real balances, r is the net nominal interest rate, π^e is expected inflation, π is actual inflation, and rr is the ex ante real rate of interest.

5. CONCLUSION

This article analyzes the economic effects of a monetary policy that responds to debt and compares those effects with ones that arise under an independent monetary policy. A principal finding is that the nominal behavior of the two calibrated economies is not very different under the two types of monetary regimes. Indeed, linear regression analysis is unable to distinguish between the two economies. It would be even more difficult to distinguish econometrically between independent and dependent monetary policy if dependent monetary policy only reacted to debt at the boundaries of the debt-to-GNP ratio rather than continuously as it does in the above example. Thus, the monetarist arithmetic is not overly unpleasant.

My analysis also indicates that in explorations of the interrelationship between monetary and fiscal policy, the exact form of the stochastic processes is important. Characteristics of the processes, such as persistence, are crucial in gauging the economic effects of policy. Policy, both fiscal and monetary, need not react immediately to changes in government debt, and the timing of the reaction is important. The empirically based persistence of the process for monetary policy is one feature of the model that leads to results that differ from those of other authors.

In addition to demonstrating the importance of the stochastic processes, the results in this article also indicate that behavioral parameters play an important role in determining the different effects that occur under the two types of monetary policy. In particular, the interest elasticity of money demand is a crucial parameter. The interest elasticity employed in this article is much lower than that used in the papers of both Leeper (1991) and Aiyagari and Gertler (1985). The lower interest elasticity is partly responsible for the influence exerted by dependent monetary policy on the nominal interest rate in my model, which is contrary to the results presented in Aiyagari and Gertler's model. Furthermore, a realistic parameterization of the interest elasticity implies that the spectacular example of Sargent and Wallace is implausible. Thus, it appears that although the intertemporal considerations highlighted by Sargent and Wallace are important theoretically, they appear to be less so in practice.

REFERENCES

Aiyagari, S. Rao, and Mark Gertler. "The Backing of Government Bonds and Monetarism," *Journal of Monetary Economics,* vol. 16 (July 1985), pp. 19–44.

Barro, Robert J., and Chaipat Sahasakul. "Average Marginal Tax Rates from Social Security and the Individual Income Tax," *Journal of Business*, vol. 59 (October 1986), pp. 555–66.

Bohn, Henning. "Budget Balance through Revenue or Spending Adjustments? Some Historial Evidence for the United States," *Journal of Monetary Economics*, vol. 27 (June 1991), pp. 333–59.

Boyd, John H., III, and Michael Dotsey. "Interest Rate Rules and Nominal Determinacy," Manuscript. University of Rochester and Federal Reserve Bank of Richmond, June 1996.

Dotsey, Michael. "Some Unpleasant Supply Side Arithmetic," *Journal of Monetary Economics*, vol. 33 (June 1994), pp. 507–24.

__________. "The Demand for Currency in the United States," *Journal of Money, Credit, and Banking*, vol. 20 (February 1988), pp. 22–40.

__________, and Ching Sheng Mao. "The Effects of Fiscal Policy in a Neoclassical Growth Model," Working Paper 94–3. Richmond: Federal Reserve Bank of Richmond, February 1996.

Drazen, Allan. "Tight Money and Inflation: Further Results," *Journal of Monetary Economics*, vol. 15 (January 1985), pp. 113–20.

King, Robert G., and Charles I. Plosser. "Money, Deficits, and Inflation," *Carnegie-Rochester Conference Series on Public Policy*, vol. 22 (Spring 1985), pp. 147–96.

Leeper, Eric M. “Equilibria Under ‘Active’ and ‘Passive’ Monetary and Fiscal Policies,” *Journal of Monetary Economics*, vol. 27 (February 1991), pp. 129–47.

Liviatan, Nissan. “Tight Money and Inflation,” *Journal of Monetary Economics*, vol. 13 (January 1984), pp. 5–15.

McCallum, Bennett T. “Are Bond-financed Deficits Inflationary? A Ricardian Analysis,” *Journal of Political Economy*, vol. 92 (February 1984), pp. 123–35.

Sargent, Thomas J. “The Ends of Four Big Inflations,” in Robert E. Hall, ed., *Inflation: Causes and Effects*. Chicago: University of Chicago Press, 1982.

__________, and Neil Wallace. “Some Unpleasant Monetarist Arithmetic,” Federal Reserve Bank of Minneapolis *Quarterly Review*, vol. 5 (Fall 1981), pp. 1–17.

Woodford, Michael. “Price-Level Determinacy without Control of a Monetary Aggregate,” *Carnegie-Rochester Conference Series on Public Policy*, vol. 43 (December 1995), pp. 1–46.