FEDERAL RESERVE BANK OF ST. LOUIS

REVIEW

SEPTEMBER/OCTOBER 2002 | VOLUME 84, NUMBER 5

REVIEW

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed at our new Web address: <**research.stlouisfed.org/publications/review**>. Single-copy subscriptions are also available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8808 or 8809.

Contents

Volume 84, Number 5

Financial Stability

Presented at the Council of State Governments, Southern Legislative Conference Annual Meeting, New Orleans, Louisiana, August 4, 2002

William Poole

I am pleased to be here to address this session of the annual meeting of the Southern Legislative Conference. Since becoming president of the St. Louis Fed, I've gotten to know pretty well a good part of the 16-state region that comprises the Southern Conference. The Eighth Federal Reserve District, headquartered in St. Louis and with branches in Little Rock, Memphis, and Louisville, includes all of Arkansas and parts of Kentucky, Mississippi, Tennessee, and Missouri. (The Eighth Federal Reserve District also includes the southern portions of Illinois and Indiana.) I've traveled extensively in this region, meeting bankers, business leaders, community and university leaders, and elected officials at all levels of government. This is a region full of vitality and, I might add as an easterner for most of my life, delightful southern hospitality.

My charge today is to discuss the condition of the national and SLC state economies. There are always many elements to analyzing the economy; I've decided to concentrate on the aspect of the current environment that seems most troubling—the condition of the equity markets.

Two hundred and fifty years ago it was established wisdom that the measure of a nation's material wealth was the size of its stock of gold. Adam Smith, in his great book, *The Wealth of Nations*, published in 1776, argued that this view was dead wrong—that the true measure was the nation's output. Today, all too often, people make a similar mistake as they judge a nation's wealth by the level of its stock market. Gold was important in Smith's day, as is the stock market in our day, but not for the reasons incorporated in the established wisdom.

My purpose today is twofold—to provide some perspective on how the stock market matters and to discuss possible approaches to creating greater financial stability.

Before proceeding, I want to emphasize that the views I express here are mine and do not necessarily reflect official positions of the Federal Reserve System. I thank my colleagues at the Federal Reserve Bank of St. Louis, especially Robert Rasche and William Emmons, for their comments, but I retain full responsibility for errors.

GOODS AND CLAIMS ON GOODS

One of Smith's essential insights, as true today as in 1776, was that gold had to be viewed as a claim on goods. The reason that people valued gold was that it could be used to buy goods they wanted—food, clothing, shelter, land, and anything else available in the marketplace. From the perspective of any one individual, gold provided command over goods and therefore was a component of the individual's wealth. But from the perspective of all individuals taken together—the entire nation—command over goods depended on the supply of goods. A nation cannot, except temporarily, consume goods beyond what it produces. For a nation as a whole to enjoy a high material standard of living—to have a large command over goods—it had to produce a lot of goods. Thus Smith argued that the wealth of a nation depends on the productivity of its people, which permits it to produce a high level of output from the hours of labor devoted to production.

Nothing has changed in this regard from Smith's day. The stock market wealth of three years ago provided each person holding a share of that wealth with a command over goods that seemed, and in the aggregate was, large. It was not possible, however, for all individuals together to cash in that wealth; for all individuals together, the goods that people could buy were limited to the goods the economy could produce. Given that we live in a global economy, we can apply that statement to all the world's citizens taken together.

BUYERS AND SELLERS

Before I discuss the role of the stock market in the economy I have to get an issue out of the way—the simple fact that every share of stock sold is also one purchased. Stock market analysts who explain

---

William Poole is the president and chief executive officer of the Federal Reserve Bank of St. Louis. The author thanks colleagues at the Federal Reserve Bank of St. Louis for helpful comments, especially Robert H. Rasche, director of research, and William R. Emmons, economist in the supervision, credit and payment risk management division. This article is a slightly revised version of the original speech. The views expressed do not necessarily reflect official positions of the Federal Reserve System.

the ups and downs of stock prices in terms of investors getting into or out of the market are not making good sense. Investors as a whole cannot get into or out of the market. An effort of investors to get out of the market depresses stock prices sufficiently that other investors are persuaded to buy. Of course, the number of shares of stock outstanding does change over time through bankruptcies, company share repurchases that retire stock, and new issues that add to the total outstanding. These factors are of trivial importance for the number of shares outstanding day by day.

Because shares sold equals shares purchased, all investors taken together cannot convert claims on wealth into goods. If one investor sells stock for the purpose of using the proceeds to buy, say, a new car, then some other investor must forego spending on goods in order to buy the shares that the first investor is selling. The effect of share prices on the economy is necessarily indirect.

Economists emphasize two mechanisms through which share prices affect the economy. One is that in a rising market companies can more easily raise funds to devote to building new factories or buying new capital equipment. Thus the level of stock prices affects the cost of capital, which in turn affects the rate of business investment in physical capital. A second mechanism is the effect of wealth on household consumption. When wealth is high, households tend to spend more of their current income, because they see less need to save for the future. When wealth declines, households tend to consume less and to save more. Thus the level of the stock market can affect households' demand for cars, TVs, vacation travel, and all the other things people spend their income on. It is important, however, to think about the wealth effect in terms of total household wealth, which includes the value of bonds and real estate as well as common stock. Finally, the evidence suggests that the wealth effect is spread out over time and is small relative to the effect of household income.

In the short run, stock market fluctuations are far, far larger than fluctuations in the nation's production, which we measure by the inflation-adjusted gross domestic product (GDP). For example, over the four quarters ending with the second quarter of this year, real GDP rose by 2.1 percent. Over the same period, the S&P 500 stock index was down 16 percent. Relative to the stock market, real GDP is so steady that we can for many purposes think of GDP as being fixed in the short run.

Given that GDP is very steady compared with the stock market, the behavior of stock prices primarily affects who gets how much of GDP rather than the total of GDP itself in the short run. If you are lucky enough to sell stock at the peak, you get more; if you are unfortunate enough to sell at the bottom, you get less. In either case, the buyer of the shares you sell is getting either less or more, the necessary mirror image of what you are getting through the accident of your timing of stock sales.

This redistribution of who gets what sometimes makes people angry, and they have good reason to be angry if the redistribution reflects market manipulation of some sort. This is one of the reasons that reforms to reduce the likelihood of market manipulation effected through accounting fraud and other means is so important. But I do want to point out that much of the redistribution between stock market winners and losers reflects outcomes that are somewhat similar to those of a lottery. No one is forced to buy a lottery ticket, and those who do should not believe that the redistribution of wealth from lottery losers to lottery winners is unfair in any respect, provided that the selection of the winners is not manipulated in any way.

Every serious student of the stock market knows that the track record of presumed expert stock pickers is not consistently better than pure random stock selection. I'm not looking to drum up hundreds of angry e-mail messages from investment professionals, and so let me add that I believe that investment professionals have a lot to offer. It is just that their clients should not believe that their investment services include reliable strategies to consistently pick stocks that will outperform the overall market and consistently identify the right times to buy and sell.

WHY THE STOCK MARKET MATTERS

When Adam Smith argued that gold was not the right measure of a nation's wealth, he was not saying that gold was irrelevant to a nation's prosperity. In his day, the monetary system was based on gold, and monetary instability clearly had negative effects on the economy. Today, the monetary system is not based on gold, and for this reason gold has little macroeconomic significance. The stock market, though not itself an adequate measure of a nation's wealth, has great importance. The market's effect on business investment and household spending on consumption goods is only part of the story.

Let me zero in on a matter of great concern to many families today. In recent years millions of

people have placed their retirement savings in the stock market. Those who placed a high fraction of their assets in certain stocks have seen their retirement dreams and their financial security disappear in the bear market underway since early 2000.

Those stock market losses could not have occurred if the market did not exhibit such large fluctuations. Suppose, hypothetically, that stock prices grew consistently along a smooth path. Take a stock market chart from 1950 to today and draw a smooth line between the starting and ending points. If stock prices grew smoothly along such a path, all the promise of rapid gains would be absent, as would all the anguish of having asset values disappear. Each stock market investor would have a high degree of certainty about his or her financial condition during retirement years.

Would investors in fact confine themselves to such stable and predictable investments? I suspect not. Indeed, I am quite certain that many would pursue strategies they believed would yield higher returns. After all, investors who went heavily into the stock market several years ago did have alternatives that were highly stable and predictable, such as government bonds, and they chose not to confine themselves to those safe havens. So I'm not sure that creating a stable stock market, if we knew how to do it, would be successful in stabilizing the retirement prospects of many people.

If the stock market does not measure the nation's wealth, what does it measure and why does it fluctuate so much? The price of a company's stock reflects market expectations about the future earnings of the company—the stock price is the present discounted value of the expected future income stream. For all companies taken together, those expectations therefore concern the country's *future* output and not its current output. Expectations are changeable because the future is uncertain and because they may be influenced by waves of optimism or pessimism. Those expectations do affect current household and business behavior, but they are far from the only determinants.

Some decry what they see as the irrational fluctuations in the stock market reflecting, they believe, expectations that get carried away on the upside or downside. I myself do not believe that it is at all easy to identify expectations that are irrational. We live in a nation that is generally exuberant about future possibilities. To my taste, we are fortunate to live in a society that nurtures invention. Our risk-taking mentality has two sides to it. On the one hand is the entrepreneurial spirit that develops new technologies and brings them to market. Many of these new technologies create astonishing improvements in our material standard of living. On the other hand is a gambling mentality that is sometimes foolish. Ahead of time, it is rarely easy to tell which bets on new businesses will work and which will not.

The importance of the stock market for the long-run performance of the economy is considerable. The longer the span of years considered, the less accurate is the assumption that GDP is roughly constant, unaffected by the behavior of the stock market. The rate of growth of GDP depends critically on the rate of productivity growth—the growth of output per hour of labor input. Productivity growth flows from innovation and entrepreneurship. A productivity growth rate of 1.5 percent per year, about what was achieved from 1968 to 1995, increases per capita GDP by 16 percent after 10 years. Since 1995, productivity growth has been about 2.5 percent per year. That rate of productivity growth increases per capita GDP by 28 percent in 10 years. There is a big difference between 16 percent and 28 percent GDP growth over the course of a decade.

Productivity growth depends on many things: One of those things is the efficiency with which the economy allocates investment, which in turn depends in part on the stock market. It can be argued that the booming stock market in the late 1990s permitted telecom companies to finance investments in computer equipment and fiber optic cable that were wasteful in the sense that this capital, even today, several years after being put in place, is not generating output and income. We would have had higher current output if the investment had gone in some other direction. From the standpoint of this particular story, the economy's productivity was damaged and not enhanced by the stock market boom in telecom shares. But the telecom mistake was not obvious at the time it occurred. If it had been completely obvious, it would not have happened. Investment mistakes are an inevitable part of a dynamic economy. We want a stock market that is receptive to new enterprises and does the best job possible in sending capital toward the most promising endeavors.

PUBLIC POLICIES TO PROMOTE FINANCIAL STABILITY

There is no realistic prospect of devising public policies that will yield stock prices that are always

"right." The future is always uncertain. New technologies are inherently experimental—some will work and others will not. From a broader perspective, the new enterprises that fail are not signs of societal failure. A business community that never fails is one that never tries.

Still, we certainly want to avoid public policies that permit, or encourage, avoidable mistakes. The current debate over accounting principles is very healthy. Penalties for fraudulent accounting and increased enforcement efforts will yield substantial societal benefits. I say "societal" and not just "economic" because a market economy that is fair, and widely perceived as fair, has benefits far beyond a higher material standard of living.

We will come out the other side of our current experience with accounting irregularities in a much stronger position than we entered it. Corporate boards, senior management, and audit firms will not take risks on accounting issues lightly. The combination of government action and market discipline has brought some prominent and long-established firms down quickly, and everyone involved in corporate governance will remember these events for a long time. The fate of Arthur Andersen, Enron, WorldCom, and other firms illustrates that the United States does have mechanisms—both governmental and market-based—to impose lasting economic reforms. Consider some other examples.

Bank failures in the 1930s led to deposit insurance. That reform contributed greatly to improved banking stability, but it turned out to have a flaw. The consequence of an inadequate regulatory system was the failure of the Federal Savings and Loan Insurance Corporation, as scores of insured savings and loan associations failed. To make good on the deposit insurance guarantee, the cost to the taxpayers in the early 1990s was in the neighborhood of $150 billion. But we learned a lesson. Regulatory requirements were strengthened; the most important of these, in my opinion, was much more rigorous enforcement of capital requirements for insured depository institutions.

We should not underestimate the contribution of this reform for improving financial stability. Failures of depository institutions in the late 1980s and early 1990s restricted the availability of credit to many borrowers, especially those that had traditionally relied on banks and S&Ls. The credit restriction was one of the reasons the economy recovered slowly from the 1990-91 recession. In contrast, last year's recession was relatively mild in part because the banking system was stable and able to lend to reasonable business risks. The stability of the banking system certainly helped the economy cope with recession.

One more example, though a smaller one: When the Penn-Central Railroad declared bankruptcy in 1970, the commercial paper market was disrupted as investors wondered what other firms were also suspect. The suspicion was in many ways a small-scale version of what we are seeing today. Until June 1, Penn-Central commercial paper was rated highly, and the company's bankruptcy on June 21 was a shock. Investors refused to roll over commercial paper of many highly rated companies because they were no longer sure what the ratings meant. Since that experience, companies have routinely arranged back-up lines of credit at banks, which they can rely on should the commercial paper market turn unreceptive. That change in business practice prevented any recurrence of the generalized disruption of the commercial paper market that we witnessed in 1970.

LOOKING AHEAD

It is easy today to look back and wish that somebody, somehow, had done more to improve accounting and audit practice. Similarly, it was easy to look back in 1990 and wish that somebody, somehow, had done more to strengthen regulation of S&Ls, to prevent the loss of $150 billion of taxpayer funds. What can we do right now to look ahead, to see what vulnerabilities we might face, and to do something in advance to ensure that some new source of financial instability does not bite us?

Periods of great market instability arise when three conditions are met. First, something happens that has widespread significance—is large enough to matter to lots of people. Second, the triggering event is a surprise; ordinarily, events long anticipated are not a problem because corrective action occurs before problems arise. Third, substantial uncertainty clouds resolution of the problem. It is especially difficult for investors to know what to do when the government's response to an unfolding situation is highly uncertain.

Let me propose two vulnerabilities we face that really need to be examined carefully. One is familiar to everyone—the state of the Social Security and Medicare systems. The issue certainly meets two of my three criteria. The potential problem is huge and there is great uncertainty about what the government

will do. Even though the problem is not a surprise in one sense, it could quickly turn into one. The fact is that a change in economic conditions could quickly increase the estimated size of the problem and move forward the time when the problem would become acute.

If the nation finds itself in a period of financial instability because of an unexpected and rapid escalation of the financial problems faced by Social Security and Medicare, we will look back and wonder why, with the vulnerability known for so long, nothing was done to reduce it. The nation has time to act, but disagreement on what should be done has led to a stalemate. Maintaining financial stability requires a willingness to find some way to engineer a compromise to reduce the nation's vulnerability that a financial crisis will some day flow from Social Security and Medicare.

The second vulnerability I would like to see more widely discussed concerns government-sponsored enterprises, or GSEs. The GSEs include Fannie Mae, Freddie Mac, the Federal Home Loan Bank System, and a number of smaller entities. The GSEs meet all three of my criteria for the potential of creating financial instability.

First, the GSEs are certainly large. In the United States today, GSE securities and government-related mortgage pool securities outstanding, excluding deposits, exceed the total outstanding securities issued by *all*—I repeat, *all*—other private financial sector firms taken together. Fannie Mae and Freddie Mac alone, as of last December 31, had securities outstanding of $1.3 trillion and had guaranteed another $1.8 trillion of mortgage-backed securities (MBS). Looked at another way, the total of GSE direct and guaranteed debt is 40 percent *larger* than the federal government's debt. That debt, which we loosely call the "national debt," has, of course, been a matter of considerable discussion in recent years in the debates about federal deficits and surpluses.

Second, although financial experts understand the vulnerability, my judgment is that too few in the markets and in government understand the issues. Consequently, if there is ever a problem, it will take many by surprise.

Third, there is tremendous ambiguity about the status of the GSEs. The market prices GSE debt as if there is a federal guarantee, or a high probability of a guarantee, standing behind the debt. Yet, there is no explicit guarantee in the law.

No one should underestimate the potential importance of the ambiguity over the financial status of the GSEs. It is not sufficient for any single GSE to argue that its own financial condition is sound. If one GSE comes under a cloud, others may also. That has been our experience again and again. It is the process economists call "contagion" whereby uninvolved or innocent firms are affected because the market has difficulty distinguishing solid firms from those at risk.

Perhaps the most famous example of contagion in U.S history is the series of bank runs in the early 1930s. Good and bad banks alike were affected. For another example, in 1970 the Penn-Central bankruptcy affected the entire commercial paper market, as investors did not know which commercial paper issuers were in fact prime credits and which, though rated prime, were not. This year, accounting problems identified in a few firms have raised questions in investors' minds about almost all firms. We may believe that only one firm in twenty, or in fifty, has suspect accounts, but how do we know which firms? We don't, and therefore investors treat all firms as suspect until the accounting treatments are verified. When there is an issue of this kind, it takes a while to get everything sorted out; in the meantime, securities prices are pushed down.

In the case of the GSEs, the massive scale of their liabilities could create a massive problem in the credit markets. If the market value of GSE debt were to fall sharply, because of ambiguity about the financial soundness of GSEs and about the willingness of the federal government to backstop the debt, what would happen? I do not know, and neither does anyone else.

Like Social Security, there are different views on what, if anything, should be done about the GSEs. In the meantime, the prevailing view seems to be that a GSE debt meltdown could not occur, or could not occur soon. I do not see any immediate risk of a GSE debt problem, but am not willing to assume that in different conditions in the future one could not occur. A judgment that there is *no* potential vulnerability seems to me to be unwarranted in light of the financial history of the United States and other countries. One thing I know for sure is that *if* the problem becomes immediate and real, then dealing with it will be very difficult because the urgency will be so great.

Let me throw out for debate two steps the federal government might take. First, various aspects of federal sponsorship that the market interprets as providing an implied guarantee of GSE debt should

be withdrawn.[1] The Secretary of the Treasury has the authority to buy GSE obligations; in the case of Fannie and Freddie, the authority is up to a maximum of $2.25 billion for each firm. The GSEs could easily replace this potential source of emergency financial support with credit lines at commercial banks, following the widespread practice among issuers of commercial paper. The amount available at the discretion of the Secretary of the Treasury is far too small in any event to deal with a crisis in the GSE debt market. Eliminating the Treasury's authority to lend to the GSEs would provide a signal that the government is serious when it says that there is no government guarantee of GSE debt. Second, over a transitional period of several years, the GSEs should add to the amount of capital they hold.

Capital is critical because, when there is a crisis in the securities markets, financially strong firms can stand the pressure without lasting damage. Capital provides a cushion against mistakes and unforeseeable circumstances. With adequate capital, a firm can almost always raise emergency loans to cover its liquidity problems.

The importance of adequate capital became clear to policymakers as the S&L problems accumulated in the late 1980s. Tightening of capital standards for insured depository institutions and the administration of those requirements was a key part of the reforms put in place at that time.

Capital is important for the GSEs because their short-term obligations are large. Fannie Mae and Freddie Mac have debt obligations due within one year of about 45 percent of their debt liabilities. Any problem in the capital markets affecting these firms could become very large very quickly.

Capital on the books of Fannie and Freddie is well below the levels required of regulated depository institutions. Let me quote a paragraph from the *2001 Annual Report* of Fannie Mae, the largest single GSE.

> During 2001, Fannie Mae issued $5 billion of subordinated debt that received a rating of AA from Standard & Poor's and Aa2 from Moody's Investors Service. Fannie Mae's subordinated debt serves as a supplement to Fannie Mae's equity capital, although it is not a component of core capital. It provides a risk-absorbing layer to supplement core capital for the benefit of senior debt holders and serves as a consistent and early market signal of credit risk for investors. By the end of 2003, Fannie Mae intends to issue sufficient subordinated debt to bring the sum of total capital and outstanding subordinated debt to at least 4 percent of on-balance sheet assets, after providing adequate capital to support off-balance sheet MBS. Total capital and outstanding subordinated debt represented 3.4 percent of on-balance sheet assets at December 31, 2001. (pp. 44-45)

The capital situation at Freddie Mac is about the same as the one at Fannie Mae. The capital adequacy standards applying to these two GSEs were established by the Federal Housing Enterprises Financial Safety and Soundness Act of 1992. The core capital requirement is 2.5 percent of on-balance sheet assets and 0.45 percent of outstanding mortgage-backed securities and other off-balance sheet obligations. The off-balance sheet obligations have a capital requirement because they are guaranteed by Fannie and Freddie.

In the private sector, government securities dealers carry capital in the neighborhood of 5 percent, and other financial firms considerably more. For example, FDIC-insured commercial banks hold equity capital and subordinated debt of nearly 11 percent of total assets.

The issue with Fannie and Freddie is not one of disclosure. Their annual reports disclose quite well the high degree of complexity of their operations, and the small amount of capital they carry over that required by law. My questions are these: Given the complexity of their operations, is the capital standard in the law adequate? Why is the standard so far below that required of federally regulated banks? What will happen to the housing market if Fannie and Freddie become unstable?

---

[1] Farmer Mac, another GSE, was much in the news in recent months. An article in the *New York Times* noted that one of the advantages conferred by government sponsorship is "the ability to borrow almost as cheaply as the government does because of a perception of government backing that emanates from a single section in its charter. That provision allows the Treasury, in certain circumstances, to provide up to $1.5 billion in loans to Farmer Mac to support the guarantees the company extends on farm loans" (9 June 2002, p. 8, col. 1).

An earlier article in the *New York Times* said the following: "The boldface disclaimers [on GSE debt offerings] state that the securities are not guaranteed by and do not constitute debts or obligations of the United States government. But the warnings are roundly dismissed by the analysts who follow the issuers' stocks, the agencies that rate their senior debt and the money managers who put their commercial paper in money market funds. In interview after interview, market professionals said that even if the paper did not carry an overt government guarantee, there was an implied guarantee, which was just as good, and the government would not allow weakness in the securities to wreak havoc. That market confidence is evident in the low interest rates that the organizations have to pay investors for financing, often only half a percentage point more than what the United States Treasury pays" (21 May 2002, p. 1, col. 5).

I've been emphasizing the importance of strengthening public policy to address potential problems. Let me add one further item to be considered—whether federal tax law should continue to encourage substitution of corporate debt for equity.

In calculating income subject to tax, corporations can deduct interest paid but not dividends paid. That provision encourages corporations to issue debt instead of equity to finance expansion and acquisitions. Firms sometimes issue debt and use the proceeds to retire equity. Many corporations today pay little or no dividends at all, preferring to provide a return to shareholders through expected capital gains on the shares, which are taxed at a lower rate than dividends in the personal income tax.

There is no doubt that a high level of debt increases the risk of financial instability. Firms fail when they cannot pay their bills. When a large fraction of revenue is devoted to paying interest instead of dividends, firms are more vulnerable to failure when revenues fall. A dividend can be cut or eliminated; interest payments cannot. Does it make good sense to maintain a feature of the tax law that makes the economy more vulnerable to financial instability? The tax law could be changed in a revenue-neutral way to eliminate this problem. I think we should do so.

CONCLUDING COMMENTS

The decline in the stock market since early 2000, and especially this summer, has been painful. We should not, however, think of the stock market as a direct measure of the nation's wealth. All you have to do is look at charts side by side of the stock market and GDP to realize that there is a long history of stock market fluctuations that are far larger than GDP fluctuations; moreover, the two are not all that highly correlated. I am not trying to tell you that the stock market does not matter, but I am trying to put the matter in proper perspective. From what we know, it is reasonable to expect that the economic recovery will continue and that the stock market will in time settle down.

This experience should make us think about what public policies could help to reduce the severity of market instability in the future. Reforms to accounting and corporate governance now being put in place are constructive. I've suggested some other things we should look at, particularly the Social Security and Medicare systems, the GSEs, and the corporate tax law. My list is not meant to be exhaustive, but surely has enough items for one speech. If any of these areas come back to bite us in the future, we'll know that the enemy is us.

Aviation Security and Terrorism: A Review of the Economic Issues

Cletus C. Coughlin, Jeffrey P. Cohen, and Sarosh R. Khan

> "Protecting this system demands a high level of vigilance because a single lapse in aviation security can result in hundreds of deaths, destroy equipment worth hundreds of millions of dollars, and have immeasurable negative impacts on the economy and the public's confidence in air travel."
> —Gerald L. Dillingham, United States General Accounting Office, in testimony before the Subcommittee on Aviation, Committee on Commerce, Science, and Transportation, U.S. Senate, April 6, 2000

The terrorist attacks exploiting weaknesses in U.S. aviation security on September 11, 2001, did indeed produce the catastrophic results identified in the prophetic testimony cited above.[1,2] Immediately after the attacks, security issues rose to paramount importance in the nation's policy agenda.[3] Despite general agreement on what aviation security entails and the goals of an aviation security system, public controversy abounds on how to regulate and provide this important activity.

If airplanes and passengers, as well as property and people on the ground, are to be protected, potential perpetrators of aviation terrorism must be prevented from breaching security checkpoints and gaining access to "secure" airport areas and to aircraft. Given the interconnectedness of the air transportation system, a sufficiently high level of security must be provided throughout the entire system. Flexibility to respond quickly to new information about aviation security threats is a must. Moreover, incentives must be offered to both the regulators and security providers so that aviation security improvements can be devised and implemented. At the same time, however, the costs associated with providing security must be incorporated in the decisionmaking process and weighed against the benefits.

In this paper we examine the economic issues relevant to airline and airport security in the United States, a topic that has received little attention from economists. Understanding the key economic issues is crucial in evaluating the various methods of regulating and providing aviation security and for appraising the conflicting positions over the appropriate scope of governmental involvement in this effort.

We begin our examination of the economics of aviation security by highlighting the key features of the airline industry, one of which is its network structure. As a result, security at one airport can affect security elsewhere—an example of a network externality.[4] Next, we use elementary economics to show that unregulated private markets will likely provide too little aviation security, which sets the stage for an examination of the alternatives for regulating and providing aviation security. We review the key features of the recently passed Aviation and Transportation Security Act and the characteristics of the resulting security policy. A summary of our major points completes the paper.

Cletus C. Coughlin is deputy director of research at the Federal Reserve Bank of St. Louis, Jeffrey P. Cohen is an assistant professor of economics at the Barney School of Business of the University of Hartford, and Sarosh R. Khan is a research associate at the Federal Reserve Bank of St. Louis. The authors thank Steven A. Morrison and Eran Segev for helpful comments.

[1] Four planes were hijacked by 19 terrorists on September 11, 2001. Two of the flights—American Airlines flight AA 11 and United Airlines flight UA 175—departed from Boston's Logan International Airport. The former flight crashed into the north tower and the latter into the south tower of the World Trade Center. The third flight—American Airlines flight AA 77—departed from Washington's Dulles International Airport and ultimately crashed into the western side of the Pentagon. The fourth flight—United Airlines flight UA 93—departed from Newark International Airport. Following passenger actions against the hijackers, it crashed in Stony Creek Township in Pennsylvania. The hijackings led to the deaths of more than 3,000 people, including all the passengers and crew on the four flights.

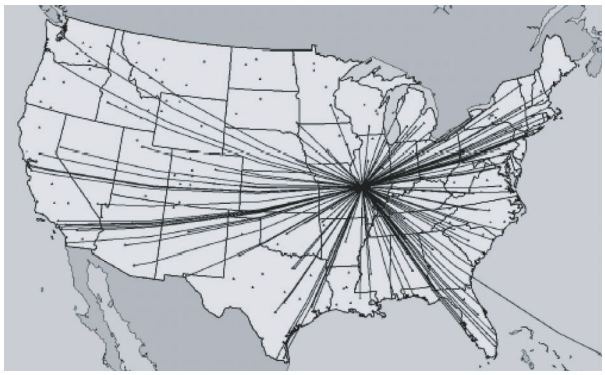
[2] We distinguish between aviation *security* and aviation *safety*. Aviation security issues require a perpetrator whose malicious intent is to advance his/her interests or that of a group, quite possibly by destroying lives and/or property. Aviation safety issues arise because of accidents due to human errors and mechanical failures.

[3] Aviation security is part of the larger issue of transportation security, which, in turn, is part of homeland security. Security policies in the United States, as well as elsewhere, have effects throughout the world. See Flynn (2000) for a recommendation that U.S. transportation policymakers pay increased attention to U.S. vulnerabilities and Flynn (2002) for a discussion of the globalization issues associated with security policies.

[4] An externality, also termed a spillover, is said to exist when either the consumption or production activity of one consumer/firm affects directly either the utility or production activity of an external party. In other words, some benefits or costs are experienced by a party that is not part of a specific consumption or production decision. The crucial economic feature of an externality is that its benefits or costs are not reflected in market prices.

Figure 1

The Lambert St. Louis International Airport Hub and Spoke System as of 12/12/01



SOURCE: U.S. Department of Transportation, Bureau of Transportation Statistics.

OVERVIEW OF THE AIRLINE INDUSTRY[5]

Prior to 9/11 the air transportation sector accounted for approximately 1 percent of U.S. employment. In 2000 there were 14 "major" certified carriers in the U.S. airline industry. Total employment (including both full time and part time) in the major carrier group was about 672,000. The major passenger carriers with the highest number of total employees were American and United, each with slightly over 100,000. Although our primary focus is on the passenger carriers, freight transport is a significant factor for several reasons. A security breach at any one airport will undoubtedly affect the smooth movement of freight through the network as well. Federal Express, one of the 14 major carriers, employed more workers than either American or United. Freight revenues overall comprise about 10 percent of total operating revenues for the major carriers, with operating revenues exceeding $20 million for each carrier. Finally, the recently passed legislation states that cargo as well as passengers will need to be screened.

Airports and Airlines: The Hub and Spoke System

Airports are a crucial component of the physical infrastructure for the airline industry. The United States has over 18,000 airports, 3,304 of which are eligible to receive federal funding. Approximately 430 airports are designated as "primary" airports by the Federal Aviation Administration (FAA). These primary airports handle virtually all the scheduled passenger service in the United States.

Subsequent to the deregulation of the airline industry that was propelled by legislation in 1978, many of the major U.S. airlines developed a "hub and spoke" system. With this structure, passengers on airline flights from various remote airports (the nodes on the spokes) converge on a single airport (the hub). After providing sufficient time for passengers to make their connections by changing planes, they depart for their final destinations. This interconnectedness is apparent in Figure 1, which shows the routes connected with the St. Louis hub. The picture would look similar for other hub cities.[6]

This hub and spoke system leads to interdependencies that give rise to several possible externalities. Namely, delays at one node often cause additional delays throughout the entire system. Thus, delays through one particular city due to security breaches can cause further delays at other nodes. For example, after a recent security breach at Atlanta's Hartsfield International Airport, an article in the *Atlanta Journal-Constitution* reported: "Hundreds of flights around the country were canceled or delayed…[and] dozens of planes heading to Atlanta were diverted to other airports."[7] Thus, by reducing travel delays throughout the system, improvements in security screening at a single airport can be viewed as a good (technically, a service) with spillover benefits.[8] Furthermore, security improvements at one node in the network can result in an increased feeling of safety perceived by passengers at other nodes. In fact, this additional safety can accrue to those who are not even traveling, such as individuals who work in high rise office buildings or in any other potential target of an airline terrorist attack.

9/11 and Airline Passenger Travel

The events of 9/11 curtailed airline travel in various ways. First, these events reduced the demand

[5] See O'Connor (2001) for a more comprehensive discussion of the economics of the airline industry.

[6] See Shy (2001) for a theoretical exposition on the development of the hub and spoke system.

[7] See Hansen and Tamman (2001, p. A.1).

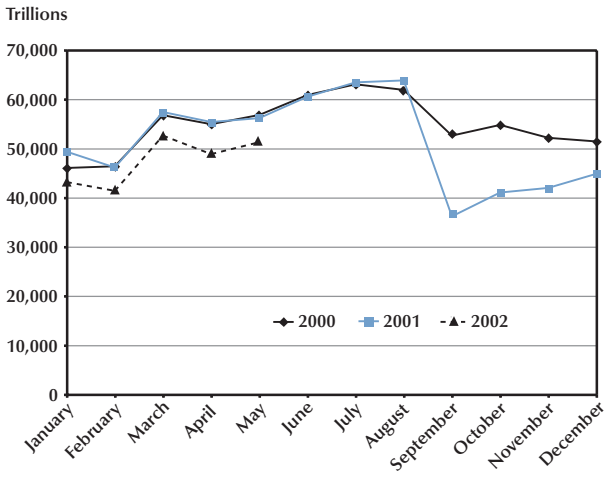
[8] Despite focusing most of our discussion on passengers, we recognize that rapid deliveries of freight—U.S. mail, checks for the Federal Reserve System, transplant organs, automobile parts, etc.—have important economic effects.

for air travel as a result of the increased concerns about safety. Second, these events reduced air travel by exacerbating the mild recession that began in March 2001. Third, the cost of travel was effectively increased because of the necessity of arriving earlier for departures, the increased frequency of delays resulting from security breaches, and new security surcharges. The result was substantially less air travel for both work and leisure purposes.

Figure 2 illustrates the dramatic drop in air passenger travel. In terms of revenue passenger miles, traffic during September 2001 declined more than 30 percent from the previous September. Despite some recovery during the fourth quarter of 2001, revenue passenger miles were down 15 percent year-over-year in December 2001. For the first five months in 2002, revenue passenger miles were 10 percent below the level in 2001. What is unclear is how long this shock will continue to affect passenger travel. Obviously, one of the major uncertainties is the effect of the new environment involving aviation security.[9]

Figure 2

Air Traffic: System Revenue Passenger Miles



SOURCE: Air Transport Association.

PROVIDING THE RIGHT AMOUNT OF AVIATION SECURITY—IN THEORY

In our introductory comments, we noted that unregulated private markets were unlikely to provide adequate aviation security. We can illustrate this claim in a relatively straightforward manner by using a supply and demand diagram. Assume that, similar to the case in the United States prior to the events of 9/11, airlines are ultimately responsible for aviation security. Assume further that consumers of airline services have a demand for this type of security, which is admittedly difficult to measure, that can be represented by the demand curve, D_p, in Figure 3.[10] The negative slope of the demand curve reflects the fact that, as the price of aviation security declines, the quantity of security that consumers desire increases. This demand curve reflects the marginal private benefits of aviation security. The supply curve for aviation security is represented by S_p in Figure 1. The positive slope indicates that increases in security can be provided only by incurring higher per-unit costs, which reflects the notion of increasing opportunity costs. The intersection of these curves generates the quantity of this good, Q_p, that is likely to be provided in equilibrium by private markets. This quantity, however, is unlikely to be the optimal (or efficient) amount of aviation security.[11]

The primary reason for underprovision in this example is that there are likely to be important benefits from aviation security that extend beyond the passengers who are on a flight. An especially gruesome example was provided by the events of 9/11. Occupants of high-rise buildings as well as those occupying other potential targets for terrorist acts (e.g., nuclear power plants and government buildings) can benefit from aviation security and, in fact, the benefits can extend beyond those individuals to their families and much further. Economists refer to this scenario as a positive externality. If positive externalities, also termed spillover benefits, exist, then the *social* demand for aviation security

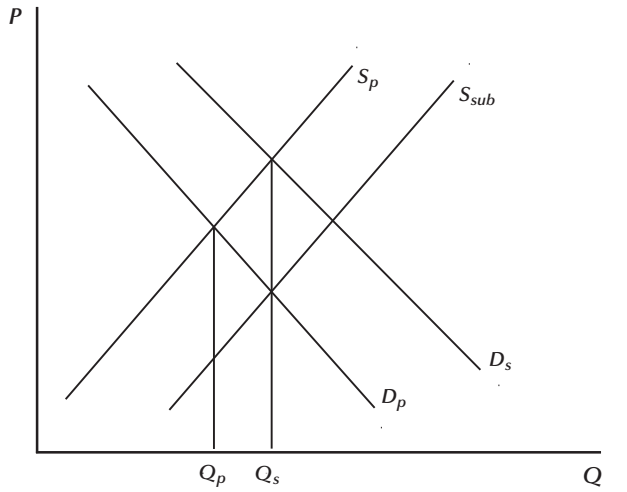
[9] A number of incidents since 9/11 have increased the fear of flying for some people. On November 4, 2001, screeners at O'Hare International Airport let a passenger with seven knives, a stun gun, and pepper spray pass through a checkpoint. On December 22, 2001, Richard Reid boarded an American Airlines flight at Charles deGaulle Airport in Paris. During the Miami-bound flight, he was overpowered by flight attendants and passengers as he tried to ignite the explosives contained in his shoes. See McTague (2002) for additional examples.

[10] Aviation security is simply one of the many attributes of air transportation service. As Moses and Savage (1990) stressed with respect to aviation safety, aviation security is not easily measured. For our purposes, we assume that a well-defined measure for safety exists that can be thought of in the following way: the smaller the probability that an airline flight will be disrupted maliciously, the larger the amount of aviation security.

[11] Based on Coase's theorem, private markets might provide the efficient quantity even when externalities exist. Provided that there are no transaction costs and given that property rights are well defined and enforceable, Coase's theorem reveals that market incentives would exist for mutually beneficial trades so that the efficient output would result. In the present case, the conditions for the Coase theorem are unlikely to exist. See Cooter (1987) for a discussion of this theorem.

Figure 3

The Optimal Quantity of Aviation Security



diverges from the private demand. This social demand encompasses the private demand plus the demand of those who benefit, but are not flying.[12] This demand curve, D_s, lies above and to the right of the private demand. The intersection of this demand curve and the supply curve determines the efficient quantity of aviation security. As Figure 3 shows, this quantity, Q_s, exceeds the quantity that would be provided by private markets.[13]

An important issue here is how to induce an increase in security from Q_p to Q_s, which leads to questions about the potential role of government—government regulation, provision, and subsidies are all possibilities.[14] Figure 3 also illustrates the effect of a subsidy. A subsidy effectively lowers the cost per unit of security and, thus, can be represented by a downward (rightward) shift of the supply curve. Assuming that the optimal subsidy is provided, this new supply curve, S_{sub}, intersects D_p at the point where the quantity of security is the socially desirable amount, Q_s. However, if the optimal subsidy is not provided, then either too little or even too much security is possible.

AVIATION SECURITY PRIOR TO 9/11

Historically, aviation security has been provided by three main partners: airlines, airports, and the FAA. Generally speaking, providing security has been the responsibility of air carriers and airports. Government, via the FAA, performed primarily a regulatory role.

The airlines were responsible for passenger and baggage screening, both carry-on and checked. The usual practice was for airlines to contract with private companies who provided trained screeners at security checkpoints. The airlines were also responsible for security from the screening checkpoints to the aircraft. Airports were responsible for law enforcement and general security in the airport vicinity, including exterior areas, parking areas, the airport perimeter, and interior areas up to the security checkpoints. The airports also hired law enforcement officers for the security checkpoints. The FAA was responsible for providing threat information; establishing security policies, regulations, and protocols; conducting security audits of airlines and airports; supporting research and development of security technology; and overseeing the installation of security equipment and devices in airports.

Aviation Security Issues

Even without factoring in the unpredictable nature of terrorism, the size of the U.S. air transportation system and the differences among airlines and airports suggest that providing aviation security is a complex and difficult task. Studies and legislation throughout the 1990s identified problems with aviation security and attempted to improve it.[15] The bombing of Pan Am Flight 103 led to the passage of the Aviation Security Improvement Act of 1990. This legislation raised employment, education, and training standards for screeners and other airport security personnel. In 1996, the crash of TWA Flight 800 led to the creation of the White House Commission on Aviation Safety and Security. This group recommended the use of new screening technologies and equipment as well as the development of

[12] Using standard terminology, the marginal social benefit, D_s, equals the marginal private benefit, D_p, plus the marginal external benefit. See Besanko and Braeutigam (2002, p. 744) for a more detailed illustration of the optimal subsidy when a positive externality exists.

[13] Despite the likelihood that private markets will underprovide aviation security, it is possible that private markets might overprovide it. Aviation security is not easily observed by consumers. Because of an information imperfection, consumers might overestimate the security threat. As a result, the demand curve might be too far to the right, leading to an excessive provision of security.

[14] Note that in our illustration the private costs of providing security include all the costs of providing security. Thus, the private costs are equal to the social costs. In a later example, we focus on how externalities associated with the network of airline transportation affect the supply of aviation security.

[15] See U.S. General Accounting Office (2000a).

uniform performance standards for training and testing screeners. Congress also passed legislation—the Federal Aviation Reauthorization Act of 1996 and the Omnibus Consolidated Appropriations Act of 1997—that provided funding for implementing many of the Commission's recommendations. For example, over the four years prior to 2000, Congress provided the FAA with $1 billion for security. Roughly one-third of this funding was for the purchase and deployment of security equipment at airports. Finally, the Airport Security Improvement Act of 2000 required additional actions to improve aviation security.

The preceding studies and legislation highlighted numerous specific problems with aviation security. Problems existed in three major areas: aviation computer security; access to aircraft, airfields, and other facilities; and the detection of dangerous objects.

With respect to aviation computer security, two major problems were well known. One problem involved the physical security at facilities housing air traffic control systems. A General Accounting Office (GAO) study (2000b) reported in 1998 that most facilities (87 of 90) had not performed threat analyses for the air traffic control systems in the five years prior to the review. A second problem involved the management of security for operational computer systems. As of December 1999, the FAA was violating its own security requirements by failing to conduct background searches on contractor employees who were reviewing and repairing critical computer system software. These employees possess critical knowledge that could prove to be very useful for computer hackers. If hackers were to penetrate the air traffic control system, they could attack the computer systems used to communicate with and control aircraft. It is not hard to imagine the physical and economic problems of a successful attack.

With respect to access to aircraft, airfields, and other facilities, controls for limiting access to secure areas had not worked as intended. For example, the results of tests during 1998 and 1999 revealed that the Inspector General's staff of the Department of Transportation successfully gained access to secure areas 68 percent of the time. These results stimulated improvements; however, additional testing between December 1999 and March 2000 revealed a rate of unlawful access of 30 percent.

The problem area that has attracted the most attention involves the detection of dangerous objects. An increase in hijackings prior to 1972 stimulated the development of passenger-screening requirements. The goal was to identify passengers carrying metallic weapons that could be used to hijack an airplane. With respect to passenger screening, personnel issues have received the most attention because screeners are not adequately detecting dangerous objects. Three reasons have been provided for this poor performance: inattention to training, high turnover, and low pay.

The previously cited GAO report revealed that the FAA was two years behind schedule in issuing a regulation implementing a congressionally mandated requirement to certify screening companies and improve the training and testing of screeners. All passengers and their carry-on baggage must be checked for weapons, explosives, or other dangerous articles that could pose a threat to the safety of the aircraft or those who board it. Until recent legislation was enacted, the FAA and air carriers shared this responsibility. The FAA set the screening regulations and established the basic standards for the screeners, the equipment, and the procedures to be used, while the air carriers were responsible for screening passengers and their baggage prior to their entry into secure areas or onto an aircraft. Generally, air carriers hired security companies to do the screening.

Concerns about the effectiveness of screeners have existed for many years. A GAO report (2000a) noted that, in 1978, screeners were not detecting 13 percent of potentially dangerous objects that FAA agents carried through checkpoints during tests. In 1987, tests revealed that 20 percent of potentially dangerous objects were passing undetected through checkpoints. Despite features of the Federal Aviation Reauthorization Act of 1996 that attempted to increase the effectiveness of screeners and the screening process, recent testimony by a GAO official (2000b) stated that the performance of screeners remained a problem. Based on the FAA's test results, which cannot be released to the public, the GAO official concluded that screeners' ability to detect dangerous objects was not improving and, in some cases, was deteriorating.

High turnover of airport security personnel is a well-known problem. From May 1998 through April 1999, turnover averaged 126 percent at 19 large airports. Skilled and experienced screeners are rare. High turnover is attributed to low wages (frequently near minimum wage), low benefits, and job stress. With respect to wages, the GAO noted that starting wages at airport fast-food restaurants frequently exceeded those of screeners.

In addition, there are some human factors associated with screening that contribute to poor performance. Screening requires repetitive tasks as well as intense monitoring for the very rare event when a dangerous object might be observed. To improve performance, the FAA began a number of programs, including establishing a threat image projection system to keep screeners alert and to monitor their performance; a screening company certification program; and screener selection tests, computer-based training, and readiness tests. However, the GAO found that the FAA's implementation was behind schedule.

The poor performance in the United States led the GAO to study screening practices in five other countries. They found lower turnover as well as differences in four areas. First, screening operations tend to be more stringent. Second, screeners' qualifications are more extensive. Third, screeners receive better pay and benefits. Fourth, responsibility usually resides with the airport's management authority or the government. Of 102 other countries with international airports, only Canada and Bermuda place responsibility with air carriers. Unfortunately, little information is available on performance; however, one joint test with another country revealed that the other country's screeners detected twice as many objects as the screeners in the United States.

Technology Issues

In addition to the personnel issues involved in detecting dangerous objects, there are technology issues. The technical performance of existing machines, which scan for metal objects, might not be adequate to detect the numerous dangerous objects that do not contain metal. A criticism of those providing aviation security is that they have failed to utilize available technology. Atkinson (2001) argues that numerous superior information technologies could and should be applied to increase aviation security. At the same time, however, the consideration of technical solutions requires the consideration of many nontechnical issues that can affect whether the technology can be implemented successfully.

New scanning technology can do a better job than the existing machines that scan only for metal. Many security experts are pushing for the use of screening machines capable of detecting a broader range of metals and alloys, plastic explosives, and other materials.

Experts are also pushing for the increased use of biometrics. Biometrics technology uses unique biological data to identify and authenticate an individual almost instantaneously. Various biological data, such as fingerprints, facial geometry, hand geometry, retinas, and voice patterns, can provide the necessary information. Plus the technical application of biometrics to increase aviation security is reasonably straightforward. For example, after background checks, an employee, such as a pilot, could be issued a card with his unique biometric information embedded on a computer chip with encrypted software. Entrance to a secure area, such as the cockpit, would require the pilot to put his card in a slot and submit to a biometric identification process to ensure that the card and the person holding it match.

A similar procedure could be used for passengers. The screening could take place both prior to entering the gate concourses and upon entering the boarding ramp to the plane. The latter authentication would allow accurate passenger manifests in real-time. This would enable airline personnel to identify individuals who have checked in, but not boarded. A related feature of this system would allow airlines to match passengers with their luggage. Luggage for an unboarded passenger could be removed.

The use of biometrics can be extended beyond the preceding examples. For example, facial biometric systems can scan individuals in a crowd or as they pass through a security checkpoint. Within seconds, a scanned face can be compared with a database of criminals or suspected terrorists. Obviously, the creation of such a database would require the cooperation of law enforcement agencies nationally as well as internationally.

The use of sophisticated technology is not simply a technology issue. In assessing the costs and benefits of using new technology, various nontechnical issues arise that can affect whether a specific technology should be utilized. First, health issues arise because the use of a technology embedded in a machine, especially one that emits radiation, might harm some individuals. Even the (inaccurate) perception that a machine might be dangerous could create adverse economic effects for the airline industry.

Second, the use of technology requires the consideration of legal and privacy issues. The technology could violate an individual's guarantee against unreasonable searches. Even if the search is legal, some potential travelers might be deterred because they feel uncomfortable with some personal information no longer being private. Understandably, many are concerned about scans that produce images

of their bodies. According to Atkinson, the new scanning technology need not reveal physiological details that create privacy concerns for passengers.

Finally, the operation of machines raises space issues because of their size and the resulting lines of passengers. Moreover, airlines are concerned about maintaining their flight schedules and the inconveniences experienced by passengers. In certain cases, it is possible that the technology can assist airlines in meeting their schedules and increase passenger convenience. To date, it appears that the greater the security threat, the greater tolerance passengers have of inconvenient procedures that increase their security.

AVIATION SECURITY IN THE AFTERMATH OF 9/11

The events of 9/11 forced public decisionmakers to examine how aviation security was being provided and how to improve it.[16] Generally speaking, three primary options for screening passengers and controlling access to secure areas were proposed before 9/11, although shortly thereafter attention focused primarily on how to implement the third option listed below. For each option identified by the GAO (2001), an underlying assumption was that the FAA would continue to regulate screening, oversee performance, and impose penalties for poor performance. These security management and provision options are as follows:

1. continue with the responsibility assigned to air carriers but with new requirements,
2. assign the responsibility to airports, or
3. assign the responsibility to the federal government via creation of a new federal agency (for example, a new agency within the Department of Transportation) or a federal corporation (for example, a corporation similar to the Tennessee Valley Authority).

Option One

The first option is the same as the pre-9/11 arrangement with the FAA promulgating new requirements. As we highlighted previously, unregulated private markets will likely provide too little aviation security. The events of 9/11 indicated that even with regulation by the FAA, too little aviation security was being provided; however, the events do not necessarily eliminate this option.

Continuing with this option implies that this system is the best way to provide aviation security. One can argue that this option worked for a number of years. The pre-9/11 security arrangements date from the early 1970s and hijackings went down markedly after these arrangements were put in place. Obviously, the hijackings of 9/11 occurred, but it is not clear that any of the options under consideration would have prevented them. It is not clear that these hijackings would have been prevented if airport security personnel were federal employees rather than privately contracted personnel. In fact, federal rules as of 9/11 would have allowed the hijackers' knives and box cutters on board because the blades were shorter than four inches. Thus, detection might not have mattered. Nor is it clear that a federal force would prevent potential hijackers from entering secure areas any better than a private force. Moreover, in light of the GAO reports cited previously, the shortcomings in the performance of the FAA justify some caution in providing more authority to a governmental body.

One can argue that the events of 9/11 revealed only that the security threat was much greater than anticipated. Furthermore, one can argue that this underestimation of the threat was not the fault of the FAA, but rather of the intelligence community at large. Of course, apart from this failure to fully recognize the security threat, our prior discussion identifying specific security shortcomings revealed that this security management and provision option, while possibly the best, is far from ideal.

As mentioned previously, this option is utilized infrequently outside of the United States. Only 2 of 102 other countries with international airports had airlines handling the security function. The primary rationale for excluding airlines from the security function was the concern that airlines would focus unduly on lowering costs and providing passenger convenience and, therefore, shirk on providing safety.

Option Two

The second option, which excludes airlines from the security function, involves assigning the

[16] Our analysis focuses on the legislated changes in aviation security rather than the changes implemented shortly after 9/11. The latter changes have not eliminated aviation security problems. Incidents reported by McTague (2002) as well as a study conducted between September 11, 2001, and February 17, 2002, reveal the continuation of problems. Morrison (2002) reported that the Department of Transportation's inspector general found that screeners missed guns 30 percent of the time, knives 70 percent of the time, and simulated explosives 60 percent of the time. In addition, in 158 tests, undercover investigators boarded 58 aircraft at 17 of the 32 airports tested and accessed the tarmac 18 times. Thus, security was breached in 48 percent of the tests.

Table 1

A Game Theory Example of Airport Provision of Security

Airport A \ Airport B	High Security	Low Security
High Security	**$800** $800	**$735** $820
Low Security	**$820** $735	**$761** $761

NOTE: Payoffs in bold are for Airport A.

security responsibilities to airports. A simple example using game theory can be used to model the network aspects of aviation security. Assume two airports—A and B—and two levels of aviation security—high and low.[17] We can think of the high level of security as allowing air travelers to have more confidence that their flight will be safe than if a low level of security were provided. In other words, the higher level of security reduces the probability of successful terrorist attempts. Table 1 shows the hypothetical payoffs of each level of aviation security for each airport. For example, the payoffs for airports A and B when A provides low security and B provides high security are $820 for A and $735 for B.

The economics underlying the payoffs in Table 1 require some elaboration.[18] Assume that the profits (payoffs) of each airport are $1000 prior to any security expenditures or any losses stemming from successful terrorist attacks. The expense of providing a high level of security is $200, while the expense of providing a low level of security is $50. Assume further that a successful act of terrorism imposes a cost of $1300 at the airport where the act occurs. If both airports provide a high level of security, acts of terrorism are prevented. If one airport provides a high level of security and the other provides a low level, then a successful terrorist act can occur at either airport; a successful terrorist act damaging the high-security airport would have emanated from the low-security airport.[19] Assume the probability of a successful terrorist act is 0.1 at an airport providing a low level of security and that the probability is 0.05 that the successful terrorist act, whose roots can be traced to the airport providing a low level of security, occurs at the other airport.

These assumptions produce the payoffs in Table 1. In the first arrangement, assume both airports provide a high level of security; both airports then receive a payoff of $800, which is simply $1000 less the $200 expense of providing a high level of security. There are no other cost calculations for this arrangement.

In the second arrangement, assume airport A provides a high level of security and airport B provides a low level of security. The payoff for airport A is $735: Starting from $1000, this airport incurs the $200 expense of providing a high level of security and an expected loss of $65. (The latter expense is the cost of a successful terrorist act [$1300] times the probability that it occurs at airport A [0.05]). Meanwhile, the payoff for airport B is $820: Starting from $1000, this airport incurs the $50 expense of providing a low level of security and an expected loss of $130. (The latter expense is the cost of a successful terrorist attack [$1300] times the probability that it occurs at airport B [0.1]). Thus, if one airport provides a high level of security and the other airport provides a low level of security, the payoff for the first airport is $735 and the payoff for the second airport is $820.

In the third arrangement, assume both airports provide a low level of security; they would each receive a payoff of $761. Starting from $1000, each airport incurs the $50 expense of providing a low level of security as well as two expected losses. The first is the $130 loss associated with a successful terrorist act occurring due to the airport's own low level of security and the second is a $59 loss (rounded from $58.50) due to the other airport's low level of

[17] This game theory framework can be extended to a case in which three or more airports provide security, but the basic economic insights are unchanged by increasing the complexity.

[18] A similar example can be found in Kunreuther and Heal (2002); however, their focus is on airlines providing security, whereas we concentrate on a network of airports that provide security.

[19] Kunreuther and Heal (2002) refer to these cross-effects as contamination.

security. (This latter loss is calculated by multiplying $65—that is, the cost of low security at one airport resulting in costs borne by the other airport—by 0.9, which is 1 minus the probability that the successful terrorist act occurred at the airport where the terrorism emanated.)

Given the preceding payoffs, what levels of security will likely be provided by the airports? Assuming that the airports make their security decisions simultaneously without communicating directly with each other, the answer is that both will provide the low level. The reasoning is straightforward. Assume airport B thinks airport A will provide the high level. If so, then if airport B also provides the high level, the payoff for airport B is $800. If airport B provides the low level, the payoff for airport B is $820. Thus, airport B will choose the low level of security because it provides the larger payoff. What happens if airport B thinks airport A will provide the low level of security? Once again, airport B will choose to provide the low level of security because the payoff to airport B is larger with the low level of security (that is, $735 versus $761). Thus, regardless of what airport A chooses, airport B will choose the low level of security. By the same reasoning process, airport A will choose the low level of security regardless of airport B's choice.

The so-called dominant strategy is for both airports to choose the low level of security. Note that the payoff for both airports is $761 and that such a payoff is inferior to the payoff of $800 to both airports if they had both chosen to provide the high level of security. Thus, when the airports choose their security level simultaneously without coordinating their decisions, there is a high probability that they will end up with lower security throughout the network. In addition, the airports will achieve lower payoffs than if they had coordinated their security decisions and jointly provided a high level of security.[20]

Option Three

The conclusion, similar to that of the first option where airlines were responsible for security provision, is that in a world in which each airport is left to provide security on its own without governmental intervention, underprovision of aviation security is likely. Thus, regardless of whether airlines or airports provide security, a role for the federal government as a regulator should not be seen as a contentious issue. Instead, the major choice for policymakers is whether the federal government of the United States should contract out the provision of aviation security services or whether it should provide those services in-house. The former scenario entails some form of public-private partnership handling aviation security. This became the norm in Western European countries during the 1990s when countries privatized aviation security following security failures by government-run operations.[21] Under this scenario, the government sets the security standards and either assigns screening responsibilities to the airport authorities or hires firms directly. Regardless, the agent is held accountable for meeting the security standards. Under this third policy option, from the list at the beginning of this section, the government is assigned full responsibility for providing security.[22]

Economic theory highlights a number of considerations regarding this third option relative to the first two options. The theory of fiscal federalism indicates the possibility of a tradeoff between (i) accounting for an externality by having a higher level of government involvement and (ii) allowing residents in individual jurisdictions to choose the desired level of public service for their own community.[23] If the federal government were to take over the provision of security at an airport, then it would be able to account for the spillover benefits by providing a higher level of airport security. However, it might do so at the cost of preventing demand diversity from being satisfied at individual airports because

[20] The numbers underlying the example were chosen to illustrate a point. It is possible that the dominant strategy could be providing a high level of security. A Nash equilibrium is also possible. In this case, an airport's best alternative depends on the security choice of the other airport. In addition, the results can be sensitive to whether the game is played just once or is repeated.

[21] Lott views this privatization as very successful. He notes that there were 21 hijackings in European airports during the 1970s, 16 during the 1980s, and 4 during the 1990s. Overall, only 3 of these 41 hijackings originated from airports with private security. Lott's argument can be found in an article on the American Enterprise Institute's Web site: <www.aei.org/oti/oti13442.htm>.

[22] The events of 9/11 generated one other contentious issue regarding governmental involvement in the U.S. airline industry. The Air Transportation Safety and System Stabilization Act included an aid package for the airline industry totaling $15 billion—$5 billion in outright cash grants and $10 billion in government-backed loans. Most agree with compensating airlines for their losses after being shut down because of the terrorist attacks. Yet, subsidies delay the adjustment of the airlines to the new economic environment. The Air Transport Stabilization Board was created to determine who receives loan guarantees and the terms. Those with a free-market orientation object because public sector employees rather than private decisionmakers are picking the winners and the losers, while those with an interventionist orientation fear that the airlines they favor will be at a disadvantage if they do not receive their fair share.

[23] See Oates (1972) for additional discussion of fiscal federalism.

the level of security is determined by the federal government. In many instances, individual communities might prefer less security at their airports than the level chosen by the federal government.

Economic theory also highlights a number of other potential problems with assigning security responsibilities to a federal agency. First, the public agency is a monopoly supplier. Similar to any monopolist, the public agency might not be forced by competitive pressures to ensure an efficient provision of services. In addition, because of civil service restrictions, the public agency might be faced with a labor environment that precludes efficient delivery of services.[24] Moreover, public agencies are frequently characterized as being slow in adjusting to changed circumstances as well as being unlikely to innovate.[25]

Additional problems might arise because the public agency is likely judged primarily on its security record. Overprovision of aviation security is possible because government bureaucrats have an incentive to protect themselves from the damage that could result if too little security is provided. In this case, the agency will have an incentive to ignore the tradeoffs that occur between security and other attributes of air transportation services that consumers demand.[26] For example, the public agency might tend to underestimate the cost of waiting incurred by passengers when it determines whether to institute a specific security measure. Waiting is a cost that airlines are sensitive to because of their profit incentive. On the other hand, the lack of a profit incentive when security is provided by the government might lead public managers to consider extended waits as simply an unavoidable cost of travel.

The fate of the following proposal, backed by the airline industry, might prove to be a good indicator of how responsive the public agency responsible for providing aviation security is to the economic interests of the airline industry. The proposal suggests creating a category of passengers known as "trusted travelers." To avoid some security checks at the airport, these travelers would endure background checks. The trusted travelers would each receive a special identification card that would allow them to proceed through a faster security line. A major concern is ensuring that the individual carrying the identification card is the trusted traveler.[27]

The heightened security measures implemented since 9/11 have already produced some examples of what could be viewed as security considerations taking precedence over other attributes of air transportation services demanded by consumers. However, one can also argue that the following examples are simply temporary costs associated with the transition to the new security environment.[28] Between October 30, 2001, and February 4, 2002, there were 35 airport terminal evacuations. Between October 30, 2001, and December 31, 2001, a total of 1,361 flights were delayed, with a cumulative delay time of 2,173 hours. During this period, 587 planes were stopped and evacuated.[29]

On the other hand, there are arguments supporting federal government provision of aviation security. First, as highlighted previously, the federal government can account for the spillover benefits associated with the provision of aviation security in its production decision. Second, governmental provision might be preferable to privatization because, relatively speaking, the former limits the incentives of managers to reduce quality by cutting costs.[30] In other words, relative to managers in private firms, managers of a government operation have less incentive to reduce quality by cutting costs because of the relatively smaller financial gains for the public employees.

---

[24] Glaeser (2001) shows that the more labor intensive the production process, the less desirable it is to nationalize the activity. Such a result could apply to airport security firms because the searching process is labor intensive. Glaeser argues that the intuition underlying his result is that when a firm is publicly owned, the workers tend to be paid in excess of the market-clearing wage rate. Thus, firms that are labor intensive are not suitable candidates for public ownership.

[25] Lott, among others, makes these points. See < www.aei.org/oti/oti13442.htm > .

[26] Holmstrom and Milgrom (1991, 1994) show that an agent with strong incentives to pursue one objective might well shirk on other objectives.

[27] According to Branch-Brioso (2002), this proposal seeks a system similar to one used since 1998 at Ben Gurion Airport in Tel Aviv. Hand scans are used to match the traveler with the identification card. Roughly 120,000 Israeli citizens are enrolled travelers.

[28] The examples can be found in Power (2002b).

[29] The increased scrutiny of passengers by screeners has sparked a privacy debate. Privacy means different things to different individuals. In some cultures and religions, the act of removing a headcovering is considered the equivalent of a public strip search. The FAA has responded to several complaints by providing detailed guidelines on performing security checks on passengers who might consider screeners' requests too intrusive. See the FAA's Office of Civil Rights at < www.faa.gov/acr > . To complicate the matter even further, several instances of harassment and abuse have been reported, some by flight crew members, since the new security measures have taken effect. See Marks (2002) and Power (2002a) for details.

[30] Hart, Schleifer, and Vishny (1997) show that if contracts are incomplete, the private provider has a stronger incentive to improve quality and reduce costs than a government employee has. However, the private provider's incentive to reduce costs is excessive because this provider ignores the adverse effects on quality that are not contractable.

In the case of aviation security, a specific concern is that private providers hire unqualified screeners and guards to minimize their costs. These attempts to cut costs undermine aviation security throughout the air transportation network. Public provision tends to mitigate this problem. This advantage of public provision is likely more pronounced the more difficult it is to specify the quality of a service. Aviation security seems to be such a case.

THE AVIATION AND TRANSPORTATION SECURITY ACT OF 2001

The Aviation and Transportation Security Act was signed into law (Public Law 107-71) on November 19, 2001, by President Bush. The act is a comprehensive approach to increasing aviation security. The objective of the act is to create, develop, and streamline security procedures and protocols that radically reduce the chances of any security breach or violation.

The enactment of the Aviation and Transportation Security Act considerably alters the aviation security responsibilities of airlines, airports, and the federal government. In the context of the three GAO options discussed previously, this legislation is the third option. A substantial increase in the resources committed to aviation security will occur as well.

The act establishes the Transportation Security Administration (TSA) in the Department of Transportation (DOT). The TSA is to be headed by the Under Secretary of Transportation for Security. As of February 17, 2002, the TSA assumed the civil aviation security functions and responsibilities of the FAA. In addition, the legislation identifies some new aviation security responsibilities. The responsibilities of this office include coordinating and directing aviation security at all times and all domestic transportation security in case of a national emergency.

The most controversial feature of the legislation is the requirement that the Attorney General and the Secretary of Transportation develop a program that ensures the screening of all passengers and baggage for illegal and dangerous items. The Attorney General is given the responsibility to develop a work force of federal employees in accordance with the guidelines of the act. This work force, which will be implemented as workers become qualified, is expected to be fully deployed by November 19, 2002. The legislation stipulates that the screeners should be subjected to background checks and that they be U.S. citizens. The TSA is also charged with ensuring sufficient explosive detection systems to screen all checked baggage at U.S. airports by December 31, 2002.

This latter objective might prove to be especially hard to achieve, especially if passenger convenience is considered in the actions necessary to meet this objective. According to Spagat (2001), fewer than 150 luggage-scanning machines capable of detecting bombs and plastic explosives were in place at 47 U.S. airports at the end of September 2001. In addition to being costly—the initial cost is roughly $1 million plus yearly costs of $700,000 to $1 million for operation and maintenance—these machines are currently slow and inaccurate. A scanner can handle only about one planeload of luggage per hour, and false alarms sound for roughly 22 of every 100 bags. Personnel must then open and search these bags. In addition, the machines can be as long as 16 feet, which poses the challenge of fitting them into existing spaces. Finally, producers of these machines might not be able to expand production rapidly enough to meet this objective.[31]

Another change is that air marshals may be deployed on all commercial flights. While the Attorney General is responsible for developing the air marshal program, the day-to-day administration of the program would be the DOT's responsibility.

Federal law enforcement officers will also be deployed to secure all areas in the larger airports, including the perimeter. A related requirement is for the DOT to improve access control systems and equipment for secured areas.

As part of a compromise to ensure passage of the legislation, the act allows for the following program. Depending on authorization by the Under Secretary of Transportation for Security, a small number of airports may employ the services of a qualified private company for the provision of airport security for up to three years. The legislation also allows other airports to opt out of the screening program and contract with private security providers after three years, if they so desire.

The legislation also contains a number of other noteworthy features. The legislation authorizes the DOT to reimburse airports for their additional costs

[31] Spagat (2001) notes that the FAA had planned to wait until 2009 to phase-in requirements for scanning all checked bags for explosives. The events of 9/11 prompted the FAA to accelerate the phase-in to 2004.

Table 2

The Cost of the Aviation and Transportation Security Act (millions of dollars)

	2002	2003	2004	2005	2006	Total
Passenger and baggage screening	889	1,942	2,181	242	0	5,254
Air marshals	92	316	561	59	0	1,028
Airport security measures	268	582	631	63	0	1,544
Reimbursement of airport authorities	553	552	0	0	0	1,105
General aviation aircraft security	19	41	45	4	0	109
R&D chemical and biological weapons	13	22	11	11	3	60
R&D aviation security technology	39	51	50	50	50	240
Regulations and reports	2	1	0	0	0	3
Estimated total cost	1,875	3,507	3,479	429	53	9,343

SOURCE: Congressional Budget Office.

of complying with increased security measures in the aftermath of 9/11. The act expands the scope of the DOT's research and development activities related to aviation security. The act requires strengthening cockpit doors and raising the quality of screening. In addition, the act allows for the needs of small airports to be dealt with by the Attorney General's office on a case-by-case basis.

The key features of the legislation can be summarized by using a concept that economists refer to as a production function. A production function shows the relationship between output, which is aviation security, and inputs, which are productive resources. Using standard terminology, the production of aviation security requires labor, capital, and technology. The labor inputs take various forms, such as passenger and baggage screeners, law enforcement officers in airports and in airplanes, managers/administrators, and researchers. The capital inputs are items such as passenger and baggage screening machines, access control systems for secured areas, and reinforced cockpit doors. Underlying the amount of output that can be produced by combining these labor and capital inputs is the level of technology, which is the body of available knowledge concerning how to combine inputs to generate maximum output. One way to increase knowledge that contributes to the increased production of aviation security is through the research and development efforts of researchers. Frequently, this new knowledge is embodied in machines and other productive resources.

Generally speaking, the legislation increases the labor and capital inputs devoted to aviation security; however, the availability of selected labor and capital inputs could prove to be a major obstacle in the near term. In addition, the legislation assigns control of these inputs to the federal government. The major unanswered question is whether the incentive system for government employees will lead to a better system in terms of the efficient production of the desired level of aviation security than any other system. Another question, somewhat easier to answer, is how much the preceding changes might cost.

Estimated Federal Government Cost

Table 2 shows a cost estimate of $9.4 billion by the Congressional Budget Office (CBO) for the expenses of the federal government.[32] The focus is on the changes in spending that are subject to appropriation for 2002-04.[33] The funds would be used for paying expenses in the following categories: passenger and baggage screening, air marshals, airport security measures, reimbursements to airports stemming from the additional security expenses due to 9/11, general aviation aircraft security, research and development on chemical

[32] The CBO's cost estimate dated October 26, 2001, was found at < www.cbo.gov/cost.shtml > . The bill number is S. 1447.

[33] Because the appropriations will occur later, the actual expenses during 2005 and 2006 for selected categories, such as "passenger and baggage" and "air marshals," are understated substantially.

and biological weapons, and research and development on aviation security technology.[34]

Passenger and Baggage Screening. The CBO estimate assumed that the Attorney General would maintain a staff of screeners similar to the existing staff employed in the private sector and that this staff would increase to keep pace with increases in passengers on domestic flights. The existing staff in the private sector consisted of 16,200 screeners, 2,800 supervisors, and 100 managers. Based on the federal pay schedule the CBO estimated that the screeners would receive an average annual base salary of $35,500, substantially higher than the average salary of screeners in the private sector of roughly $15,000. To generate an estimate of the actual costs per screener, this average base salary was adjusted upward by benefits of 35 percent of the base as well as by overtime pay. The CBO estimates used an average salary of $52,600 for supervisors and $74,900 for managers. These salaries were adjusted for benefits identical to the screeners, but no overtime pay was anticipated.

The legislation also authorizes the Attorney General to deploy at least one law enforcement officer at each of the existing 754 airport checkpoints. Thus, at a minimum, to staff each checkpoint around the clock requires 2,262 officers. The Attorney General has the authority to deploy more officers at the 100 largest airports. The CBO estimates used an average salary of $46,500 for these officers. Benefits plus overtime increase the average cost for each officer to $73,000.

In addition to the personnel involved directly in screening and law enforcement, there are a number of other costs. First, there are expenses associated with the required administrative staff. Second, there are costs for training, testing, and auditing screeners and for performing background checks. Third, the legislation requires a senior level security officer at each airport (about 450 positions) and two ground security coordinators at each checkpoint (about 1,500 positions). Fourth, additional screening equipment must be purchased, installed, and maintained. The total costs for screening and law enforcement are estimated to be $5.3 billion.

Air Marshals. The legislation authorizes the presence of air marshals on all scheduled flights. Whether or not an air marshal would fly on all scheduled flights is to be determined by the Attorney General. The CBO assumed that an air marshal would fly on 20 percent of all flights.[35] As a result, the number of required air marshals would be 2,800. The CBO estimated an average cost per marshal, including salary, benefits, training, supervision, equipment, and other administrative expenses, of $170,000 and a total cost of $1 billion.

Airport Security Measures. The legislation authorizes a variety of measures estimated to cost $1.5 billion to increase security at airports. First, the legislation authorizes the deployment of federal law enforcement officers to secure all areas in the nation's largest airports. Second, the Secretary of Transportation is to work with small- and medium-sized airports to determine their needs. This might lead to the deployment of federal law enforcement officers in these airports as well. Third, the Secretary of Transportation is to work with airport operators to improve access control systems and equipment for secured areas.

The CBO estimates that 6,990 federal law enforcement officers would be deployed at an average cost per officer of $85,000. Each of the 120 largest commercial airports would have 50 federal law enforcement officers. On average, the smaller airports would have three federal law enforcement officers.

Reimbursement of Airports for Increased Security Costs. The legislation authorizes the Secretary of Transportation to reimburse airports for their fiscal year 2002 costs associated with complying with the 9/11-induced security measures. The costs cover additional law enforcement personnel, access-control equipment, and operating costs. Some of these upgrades will not be completed in 2002, so roughly one-half of the $1.1 billion cost will be incurred in 2003.

General Aviation Aircraft Security. The legislation requires the FAA to develop a program to search general aviation aircraft (i.e., private aircraft and charter planes) as well as screen crew members and others who might board a flight prior to takeoff. The CBO estimates the cost of this security enhancement to be $109 million for the 2002-04 period.

Research and Development. The legislation authorizes the FAA to expand research in two areas. First, the FAA is authorized to conduct research concerning chemical and biological warfare and

[34] A final category involving regulations and reports is not discussed because of its small (less than $3 million) budgetary effects.

[35] McTague (2002) argues that two air marshals should be on each commercial flight in the United States. Since the Israelis began such a program in 1986, no El Al flight has been hijacked.

to develop technologies to prevent the successful use of these weapons in planes and airports. Second, the FAA is to increase support for research and development related to all aspects of aviation security involving technology, such as detecting explosives; screening baggage, passengers, and cargo; training employees; and constructing aircraft. The FAA's support would be in the form of grants to industrial, academic, and governmental entities for promising projects. In addition, the FAA is authorized to provide research grants dealing with biometrics, longer-term airport security, and information sharing among federal agencies. In total, the estimated cost of research and development is $300 million.

Estimated Impacts on Non-Federal Governments and the Private Sector

The legislation requires numerous actions by airport operators and, depending on how the FAA and Department of Justice choose to implement other requirements in the legislation, may necessitate other actions. In the former category are requirements that airport operators use technology to detect weapons, develop security awareness programs for airport employees, and conduct background checks on employees with access to planes and secure areas. In the latter category are requirements involving security around airport perimeters, the screening of passengers at smaller airports, and the screening of personnel and supplies entering secure areas.

Generally speaking, airport operators have already taken actions to comply with FAA regulations following 9/11. The additional costs are not expected to exceed $56 million annually (in 2001 dollars). Moreover, the legislation authorizes funding for airports to cover the costs of security improvements resulting from post-9/11 requirements.

With respect to the impact on the private sector, the legislation imposes mandates affecting air carriers, commercial airplane manufacturers, persons providing training in operating aircraft, and aliens. The Department of Transportation has imposed a $2.50 fee for each passenger enplanement that will be remitted by the airlines to the federal government to pay for the federal government's costs of providing aviation security. Because air carriers would no longer be responsible for screening passengers and baggage, it is uncertain whether the net income of air carriers would rise or fall.

The bill requires commercial manufacturers to increase the security involving the doors separating the pilots from the passengers on new large aircraft as well as on new commuter aircraft. The cost of this mandate depends on the standards set by the FAA.

Finally, the legislation mandates that persons who provide aircraft training report certain information on those they train. Aliens would be required to undergo a background check from the Attorney General prior to training. The expectation is that the costs of these mandates would be small.

CONCLUSION

One unsettling conclusion following the events of 9/11 was that both the quantity and quality of aviation security, each difficult to measure, were inadequate. Quite likely both demand and supply factors underlie this conclusion. On the demand side, the catastrophic events of 9/11 increased the demand for aviation security by increasing awareness of the very real security threat that existed and likely continues to exist. Moreover, the events of 9/11 focused attention on how aviation security was being provided and regulated. This attention revealed numerous shortcomings that prompted increased scrutiny of not only how much aviation security was being provided, but also how it was being provided.

Public decisionmakers have been prompted to ensure that more resources will be devoted to providing aviation security today as well as to research and development activities that should lead to improvements in aviation security in the future. In addition, changes were made in who has the authority concerning aviation security decisions. The hope is that these changes will result in the provision of the efficient level of aviation security.

Economic theory can be used to make a strong case that the federal government play an important role in aviation security. The basic question is whether the federal role should be restricted to setting and monitoring security standards or whether the role should also include the financing and implementation of security. The most contentious change emanating from 9/11 is that the federal government has assumed responsibility from the airlines and airports for the actual provision of aviation security. Policymakers assigned the responsibility for aviation security to the federal government, primarily through the authority vested in the newly created Transportation Security Administration and the Department of Justice.

Will this substantial enlargement of governmental involvement, which is in contrast to the public-private partnerships that dominate aviation security

in Europe, be a change for the better? In theory, public provision of aviation security can adequately account for security externalities. Moreover, relative to private provision, public provision reduces the incentives to reduce quality by reducing costs. Proponents of in-house provision argue that the quality of public services delivered by government employees is superior to that delivered by private firms. This feature of public provision might be especially relevant for a service, such as aviation security, whose quality is hard to observe.

On the other hand, a public agency might not provide security services efficiently because it can operate in a more-or-less monopolistic manner. Proponents of government contracts with private suppliers argue that private firms deliver public services at a lower cost than the government does. In addition, responsiveness to the consumer is not a trademark of monopolistic markets.

Furthermore, it is also possible that a public agency with one objective might provide an excessive amount of security (and incur excessive costs) because it is likely to be judged primarily on its security record and not on all the attributes encompassed by air transportation services for consumers. If either or both situations occur, then adverse consequences would result for both consumers and suppliers of air transportation services. At this point, given the still vivid memories of 9/11, the general public is likely to prefer too much aviation security to too little. However, one cannot conclude that public provision is a panacea.

The more important question is whether public provision will be an improvement relative to the less-than-perfect pre-9/11 system for providing aviation security. It is too early to answer this difficult question.

REFERENCES

Atkinson, Robert D. "How Technology Can Help Make Air Travel Safe Again." Policy Report, Progressive Policy Institute, September 2001.

Besanko, David A. and Braeutigam, Ronald R. *Microeconomics: An Integrated Approach*. New York: Wiley, 2002.

Branch-Brioso, Karen. "U.S. Official Backs 'Trusted Traveler' Plan for Airports." *St. Louis Post-Dispatch*, 23 April 2002, p. A3.

Cooter, Robert. "Coase Theorem," in John Eatwell, Murray Milgate, and Peter Newman, eds. *The New Palgrave: A Dictionary of Economics*, Vol. 1, A to D. New York: Macmillan, 1987, pp. 457-60.

Flynn, Stephen E. "America the Vulnerable." *Foreign Affairs*, January/February 2002, *81*(1), pp. 60-74.

_____ . "Transportation Security: Agenda for the 21st Century." *TR News*, November-December 2000, pp. 3-7.

Glaeser, Edward L. "Public Ownership in the American City." Working Paper No. W8613, National Bureau of Economic Research, December 2001.

Hansen, Jane O. and Tamman, Maurice. "UGA Fan's Hunt for Camera Bag Turns World's Busiest Airport into Haltsfield." *The Atlanta Journal-Constitution*, 17 November 2001, p. A.1.

Hart, Oliver; Schleifer, Andrei and Vishny, Robert W. "The Proper Scope of Government: Theory and Application to Prisons." *Quarterly Journal of Economics*, November 1997, *112*(4), pp. 1127-61.

Holmstrom, Bengt and Milgrom, Paul. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization*, Special Issue 1991, *7*(0), pp. 24-52.

_____ . "The Firm As an Incentive System." *American Economic Review*, September 1994, *84*(4), pp. 972-91.

Kunreuther, Howard and Heal, Geoffrey. "Interdependent Security: The Case of Identical Agents." Working Paper No. W8871, National Bureau of Economic Research, April 2002.

Marks, Paul. "Flight Attendants Find Security Abusive." *Hartford Courant*, 18 February 2002.

McTague, Jim. "Wanted: Wyatt Earps." *Barron's*, 4 March 2002, pp. 21-22.

Morrison, Blake. "Tests Show No Screening Improvements Post-Sept. 11." *USA Today*, 25 March 2002, p. A4.

Moses, Leon N. and Savage, Ian. "Aviation Deregulation and Safety." *Journal of Transport Economics and Policy*, May 1990, *24*(2), pp. 171-88.

Oates, Wallace E. *Fiscal Federalism*. New York: Harcourt Brace, 1972.

O'Connor, William E. *An Introduction to Airline Economics*. New York: Praeger, 2001.

Power, Stephen. "Airport Screens Cause Conflicts with Carriers." *Wall Street Journal*, 22 February 2002a, p. A3.

_____ . "Airlines and Stadiums Have Boosted Security, but Much Remains Undone." *Wall Street Journal Online*, 8 March 2002b.

Shy, Oz. *The Economics of Network Industries*. Cambridge: Cambridge University Press, 2001.

Spagat, Elliot. "Sophisticated Bag Scanners Fail to Catch On As Airlines Complain of Delays, False Alarms." *Wall Street Journal Online*, 29 October 2001.

U.S. General Accounting Office. *Aviation Security: Long-Standing Problems Impair Airport Screeners' Performance*. Washington, DC: June 2000a.

_____ . *Aviation Security: Vulnerabilities Still Exist in the Aviation Security System*. Washington, DC: April 2000b.

_____ . *Aviation Security: Weaknesses in Airport Security and Options for Assigning Screening Operations*, Testimony before the Subcommittee on Aviation, Committee on Transportation and Infrastructure, House of Representatives. Washington, DC: 12 September 2001.

Has Japan Been Left Out in the Cold by Regional Integration?

Howard J. Wall

I. INTRODUCTION

A regional trading bloc (RTB) is a grouping of countries in which trade between members faces fewer restrictions (i.e., tariffs, quotas, nontariff barriers) than trade between a member and a nonmember. One of the most significant recent trends in international trade has been the increasing importance of RTBs, which have been growing in both number and size for over a decade.[1] Not all countries have followed the trend toward regional integration, though, and relatively little research has examined the effects of regional integration on non-integrators. This paper attempts to fill this void by estimating the effects of several major RTBs on the trade pattern of the largest non-integrator, Japan.

From the perspective of Japanese trade, the two most important RTBs are the North American and European trading blocs, whose most recent regimes are the North American Free Trade Agreement (NAFTA) and the European Economic Area (EEA). NAFTA was inaugurated in 1994, bringing Mexico into the RTB that had been in place between Canada and the United States since 1989. The EEA includes the 15 members of the European Union (EU) and the four members of the European Free Trade Area (EFTA). It was formed in 1995, although the EU and EFTA had maintained separate RTBs of varying depth and breadth since 1957 and 1960, respectively.

A third bloc, the Association of South East Asian Nations Free Trade Area (AFTA), is potentially as important for Japan as the EEA: each bloc accounts for around 16 percent of Japan's trade.[2] AFTA, however, has not been nearly as deep an integration regime, so its importance for Japan is yet to be fully realized. Frankel (1997) reports dozens of other RTBs around the world, most of which are among small countries. The most important of these from a Japanese perspective are the Australia–New Zealand Closer Economic Relations Trade Agreement (ANZCERTA), the Mercado Comùn del Sur (Mercosur), and the Andean Community.[3]

Japan has been notably reluctant to follow the trend toward regional integration, maintaining a multilateral approach through the World Trade Organization (WTO), while monitoring RTBs for any tendencies toward higher protection against nonmembers (Ministry of Economy, Trade, and Industry, 2001). Recently, though, Japan has taken a more nuanced, "multilayered" approach designed to extract the benefits of multilateral integration while avoiding many of the discriminatory consequences associated with RTBs.[4] As described by Eguchi (2001), one of the main reasons for this new approach is to establish footholds within existing RTBs to avoid some of the discriminatory tariff treatment that Japanese goods would face otherwise. This approach has led to recent bilateral discussions with Singapore, South Korea, Mexico, and Chile.

For the time being, however, Japan has bucked the regional integration trend while its trading partners have become increasingly regionally integrated: by the late 1990s, more than 60 percent of Japan's total trade (imports plus exports) was with countries that were members of the six trading blocs described above. As a consequence, there has been ongoing concern that Japan has been left at a disadvantage when its firms compete within RTB markets, and that its trade patterns have been disrupted. To date, though, I have found little research done on the extent to which RTBs have affected

Howard J. Wall is a research officer at the Federal Reserve Bank of St. Louis; this research was done while he was a visiting scholar at the Institute for Monetary and Economic Studies (IMES) of the Bank of Japan (BOJ) during the spring of 2001 and was previously published (Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies*, April 2002, *20*(2), pp. 117-34). The author thanks the IMES and the BOJ for hospitality and research resources and the seminar participants at the IMES, Hitotsubashi University, and the Research Institute of Economy, Trade, and Industry for their comments. The views expressed are the author's alone and are not necessarily those of the IMES, the BOJ, the Federal Reserve Bank of St. Louis, or the Federal Reserve System.

[1] See Frankel (1997) for excellent descriptions and histories of the various RTBs.

[2] AFTA includes Brunei, Indonesia, Malaysia, the Philippines, Singapore, and Thailand. Although a trade bloc of sorts had been in place since 1978, its coverage was extremely limited. AFTA was created in 1992 and began to be implemented in subsequent years (Frankel, 1997).

[3] ANZCERTA has been in place since 1983. Mercosur was formed in 1995 and includes Argentina, Brazil, Paraguay, and Uruguay. The Andean Community includes Bolivia, Colombia, Ecuador, Peru, and Venezuela.

[4] See Eguchi (2001) and Kojima (2001) for discussions of RTBs from Japan's perspective.

Japan, although Anderson and Snape (1994) argue that many of the concerns among non-integrating Asian countries are unwarranted.

Because this paper focuses on the effects of RTBs on a nonmember, it is in a minority of research on RTBs, as nearly all previous research has been from the perspective of member countries.[5] For member countries, the trade effects are usually categorized as trade creation (the supposed increase in trade between members) and trade diversion (the supposed decrease in imports from nonmembers). This dichotomy has dominated the discussion of the effects of RTBs since Viner (1950) first used it to establish the general welfare ambiguity of RTBs for member countries. For nonmembers, the usual presumption has been that the reduction in exports to RTB members would make them worse off.

The empirical analysis of the trade effects of RTBs has advanced little beyond the Vinerian effects, even though they were derived in a simple single-industry partial equilibrium model. As shown by Winters (1997), however, in a standard general equilibrium model, an RTB affects not only a member's imports from nonmembers, but also its exports to them. This aspect of RTBs is usually ignored in empirical studies, and when it has been considered, imports and exports are often lumped together under the extremely suspect assumption that Vinerian trade diversion applies to exports as well as imports. Further, as argued by Wall (2000) and Cheng and Wall (2001), under capital mobility an RTB will affect not just the quantities that firms produce, but also the countries in which production takes place. This has obvious implications for Japan's trade pattern for, as shown by Head and Ries (2001) and Lipsey, Ramstetter, and Blomström (2000), Japanese firms that increase their manufacturing investment overseas also tended to increase their exports.

The effects of RTBs on trade between members and nonmembers discussed in the preceding paragraphs arise even when the level of protection toward nonmembers is unaffected by integration. An emerging literature, though, has looked at how joining an RTB can change levels of protection. In Yi (2000) and Kose and Riezman (2000), if the RTB is a free trade area—in which members set their own tariffs—tariffs are lower after integration. On the other hand, a customs union—in which members have a common tariff—might lead to higher post-integration tariffs (Kose and Riezman, 2000, and Bandyopadhyay and Wall, 1999).

To summarize, even assuming that capital does not migrate in response to RTBs, it is not necessarily true that Japanese trade with an RTB's members will be reduced, despite the presumed certainty of trade diversion. Also, once you consider the mobility of capital, it is possible that some or all RTBs have led to increases in Japanese exports and/or imports. And finally, the level of protection toward nonmembers can change after integration, and the direction of this change can depend on the type of integration that is chosen. What this means is that the question posed by the title of this paper is not as straightforward as would be suggested by the simple Vinerian dichotomy. It is not possible to make reliable a priori predictions about the signs of the effects of an RTB on Japanese trade.

The rest of the paper is organized as follows. Section II briefly discusses some recent trends in Japanese trade, particularly those regarding trade with major RTBs. The empirical model is presented in Section III, and Section IV presents the empirical results. Concluding remarks are presented in Section V.

II. RECENT TRENDS IN JAPANESE TRADE

The bilateral trade data that I use come from the *World Trade Flows* dataset described in detail by Feenstra (2000). These data are the United Nations trade data recompiled by Statistics Canada to make them consistent across countries and over time. This recompiling is necessary to avoid various problems with the original data, including discrepancies between import and export reports, i.e., A's reported exports to B do not match B's reported imports from A. Unfortunately, because of the scale of the project, the data are available only with long lags, meaning that 1997 is the latest year available.

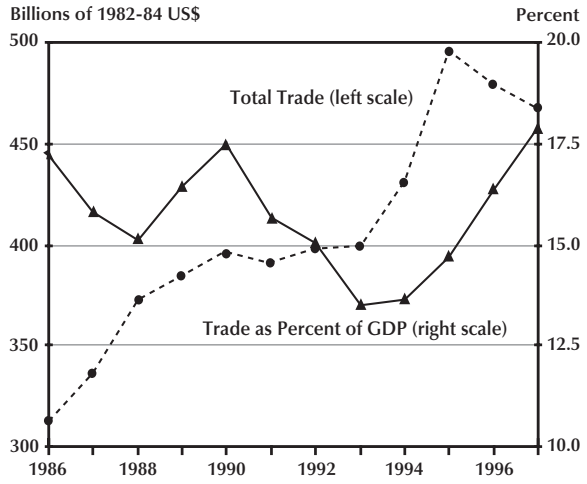
As illustrated by Figure 1, Japanese real total trade (imports plus exports) measured in U.S. dollars trended upward between 1986 and 1997, peaking in 1995 at nearly 60 percent above its 1986 level.[6] However, looking at total trade relative to the size of the Japanese economy, this consistent upward trend disappears. By 1993, total trade as a percent of GDP had fallen to 13.5 percent, having risen to nearly 17.5 percent in 1986 and 1990. By 1997,

[5] Exceptions include Winters (1997), Goto and Hamada (1998, 1999), and Winters and Chang (2000).

[6] This is at market exchange rates and the U.S. consumer price index relative to average prices for 1982-84.

Figure 1

Japan's Real Total Trade (Imports plus Exports)



SOURCE: *World Trade Flows* 1980-97 and the World Bank.

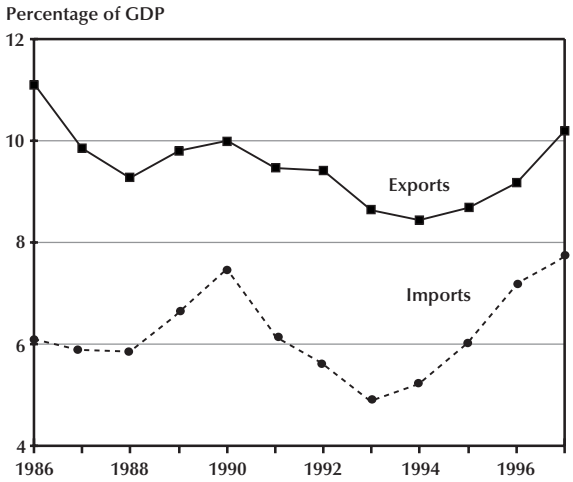
Figure 2

Japan's Real Imports and Exports

SOURCE: *World Trade Flows* 1980-97.

Figure 3

Japan's Imports and Exports Relative to GDP



SOURCE: *World Trade Flows* 1980-97 and the World Bank.

though, it had recovered and was back at a new peak for the period of close to 18 percent.

Figures 2 and 3 illustrate how export and import trends differed somewhat from each other. While real imports and exports were both significantly higher in 1997 than in 1986, real imports did not rise steadily, and even fell by more than 15 percent between 1990 and 1993. Real exports, on the other hand, rose steadily throughout the period. Also, despite the dip in the early 1990s, real growth in imports between 1986 and 1997 (82 percent) easily outstripped the growth in real exports (32 percent). Further evidence of the trend toward imports relative to exports is that exports as a percentage of GDP were actually slightly lower in 1997 than in 1986, whereas imports as a percentage of GDP were nearly one-third higher over the same period. Greaney (2001) notes that the increase in imports has not been due to an opening of Japanese markets to imports, but is related to increased imports from overseas affiliates of Japanese firms and importing by Japan-based affiliates of foreign firms.[7]

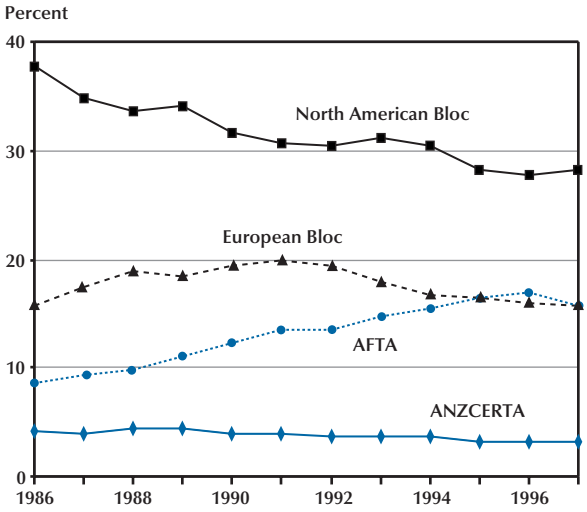
The primary concern of this paper is the geographic allocation of Japanese trade across RTBs, which, as shown by Figure 4, has seen some changes in recent years.[8] The most notable of these has been the decreasing importance of the members of the North American and European trading blocs relative to members of AFTA. The decline in North America's share extended over the whole period, although

[7] She finds that "by the late 1990s, slightly over half of Japan's imports are…provided by Japanese affiliates abroad and approximately one quarter of imports are purchases made by foreign affiliates in Japan."

[8] To construct these figures, trade in a given year is the total of trade with all countries that were members of the respective blocs in 1997.

Figure 4

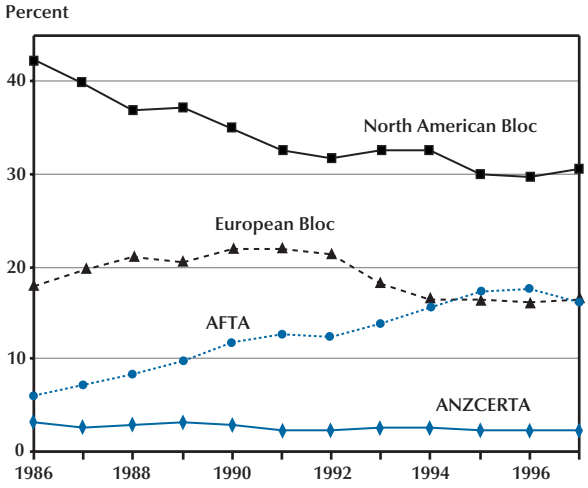
Major Blocs' Shares of Japan's Total Trade



SOURCE: *World Trade Flows* 1980-97.

Figure 5

Major Blocs' Shares of Japan's Exports

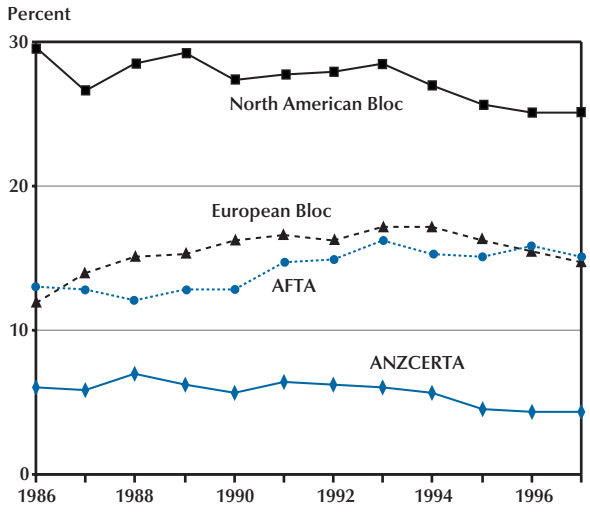


SOURCE: *World Trade Flows* 1980-97.

there is evidence that it was encouraged by regional integration. There was a noticeable drop in the North American bloc's share following integration between the United States and Canada in 1989, and again following the addition of Mexico in 1994. Similarly, for Japanese trade with the European bloc, there was a notable turning point after 1992, when

Figure 6

Major Blocs' Shares of Japan's Imports



SOURCE: *World Trade Flows* 1980-97.

the European Community deepened its integration and renamed itself the European Union.

A disaggregation of total Japanese trade into imports and exports, as in Figures 5 and 6, reveals even more about the potential effects of integration on the distribution of Japan's trade. Most of the decline in North America's share of total Japanese trade has been in its share of Japanese exports, although its import and export shares both dropped noticeably following the two stages of North American integration. For post-1992 trade, European bloc members saw their share of Japanese exports drop more sharply than their share of Japanese imports. Meanwhile, the shares for the members of AFTA trended upward for both imports and exports, although the larger increase was in the share of Japanese exports going to these countries. In 1986, AFTA members received less than half as much of Japanese exports that European bloc members received, but by 1997 they received an equal share. In contrast, the blocs' shares of Japanese imports were very similar throughout the period.

Hand in hand with the trends in the geographic components of Japan's trade are the trends in the mix of goods that Japan imports and exports (Ministry of International Trade and Industry [MITI], 1998). Over the period, Japan's imports have shifted away from raw materials and toward manufactured goods, particularly to parts of machinery and trans-

portation equipment. Much of this shift has been reflected in increased shares of these goods from East Asian countries, and a decreased share from the United States. On the export side, the share of capital goods continued to increase as it had since the 1970s, which can be attributed to the growing presence of Japanese firms with production facilities inside overseas markets (MITI, 1998).[9] Instead of producing all of the consumer durables that its firms sell in these markets, Japan exported more machines and intermediate products, assembling the consumer durables in the markets where they were sold (or were even exported from this new production base). Of course, the other side of this trend is that exports of durable consumer goods such as automobiles decreased at the same time. However, these changes have not had the geographic dimension that was apparent on the imports side.

III. THE EMPIRICAL MODEL

Although the trends in the geographic distribution of Japan's trade illustrated by Figures 4 through 6 are consistent with what one might expect following the formation of RTBs, they are only suggestive. While they suggest that Japanese trade levels have been affected adversely by North American and European integration, they should not be taken too seriously, because they control for very few of the many determinants of trade. To obtain more rigorous estimates of the effects of RTBs on Japanese trade, I use a gravity model, which recently has become the workhorse of empirical studies of international integration.

The gravity model of international trade assumes that the volume of bilateral trade can be estimated as an increasing function of the sizes of the trading economies, and a decreasing function of the geographic distance between them. In Tinbergen (1962) and Pöyhönen (1963), the sizes of the economies were measured simply by their national incomes, although, since Linnemann (1966), it has been common to add their populations or per capita incomes to the model. To control for various other factors, it has also been standard to include dummy variables to indicate when trading partners have colonial or linguistic links, are contiguous, are islands, etc. (for a survey, see Oguledo and MacPhee, 1994). For many years, the gravity model was criticized as being ad hoc, although recent theoretical justifications for it have led to its wider acceptance. Recently, for example, Deardorff (1998) showed that the gravity model can be consistent with several standard trade models.

Gravity models have been used to examine Japanese trade by Eaton and Tamura (1994, 1996), who looked at differences between Japanese and U.S. trade and direct foreign investment patterns. Alternatively, vector autoregression methods have been used by Ceglowski (1996), Daly (1998), and Nadenichek (2000) to examine, respectively, the late 1980s surge in U.S. imports from Japan, the effect of exchange rate volatility on Japanese trade, and the Japan-U.S. trade imbalance in a real business cycle model. Also, see Sawyer and Sprinkle (1997) for a survey of the empirical international economics literature as it applies to Japanese trade.

The gravity model I estimate is standard, except that, following Mátyás (1997), Bayoumi and Eichengreen (1997), and Cheng and Wall (2002), I allow the intercept to differ across trading partners. Specifically, I estimate

(1)
$$\ln(x_{ijt}) = \alpha_0 + \alpha_{ij} + \gamma \mathbf{T}'_{\mathbf{ijt}} + \beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta_3 \ln y_{it} + \beta_4 \ln y_{jt} + \delta \ln D_{ij} + \sum_{b=1}^{6} \begin{pmatrix} \theta_1^b I_{ijt}^b + \theta_2^b M_{ijt}^b + \theta_3^b X_{ijt}^b \\ + \theta_4^b MJ_{ijt}^b + \theta_5^b XJ_{ijt}^b \end{pmatrix} + \varepsilon_{ijt},$$

where x_{ijt} is real exports from country i to country j in year t, Y_{it} and Y_{jt} are their real GDPs, y_{it} and y_{jt} are their real per capita GDPs, and D_{ij} is the distance between them. In addition to these gravity variables, equation (1) also includes a vector of time dummies $\mathbf{T}_{\mathbf{ijt}}$; a trading-pair effect, α_{ij}, that is fixed over time; and 30 integration dummies.

Gravity models usually include variables to capture fixed factors that are thought to affect bilateral trade, including dummies for contiguity and a common language, among others. In equation (1), these factors are all subsumed into trading-pair fixed effects. Specifically, $\alpha_{ij} = \omega \mathbf{Z}'_{\mathbf{ij}}$, where $\mathbf{Z}_{\mathbf{ij}}$ is a vector of all of the fixed factors that make the volume of exports from i to j differ from the average, and ω is a vector of their weights. Rather than trying to specify all of the unknown number of these variables, which may or may not even be observable, I simply specify their total effect as a trading-pair-specific intercept term. Further, because α_{ij} and $\delta \ln D_{ij}$ cannot be estimated separately, I estimate their sum as a fixed effect, $\pi_{ij} = \alpha_{ij} + \delta \ln D_{ij}$, using a dummy variable for each of the trading pairs. A

[9] See Abe and Zhao (2000), Kimura (2000), and Lipsey (2000) for studies of the geographic and sectoral dimensions of Japanese direct foreign investment.

trading-pair dummy takes the value of one whenever the observation is of exports from i to j and is zero otherwise.[10]

Using fixed-effects estimation allows me to capture the influence of not only distance and other fixed variables normally specified in gravity models, but also of any number of important factors that may be difficult or impossible to quantify and, therefore, are normally excluded. As shown by Cheng and Wall (2001, 2002), excluding these variables (which is the same as restricting their effects to be the same across states and countries) results in serious estimation bias. In particular, they show how the usual estimation methods hugely overestimate the effects of RTBs on trade because they do not properly account for the possibility that there are factors responsible for both high levels of trade between members and the probability that they will join the same RTB.

Another advantage of modeling these factors as fixed effects is that it allows me to avoid the ad hoc measures of distance normally used, while still controlling for the effects of distance on trade volume. Distance in gravity models is usually simply the great circle distance between the capital cities or the geographic centers of the two countries, which is clearly unsatisfactory.

The variables of most present interest are the integration dummies. For each of the six RTBs, I specify five dummy variables to capture each of its effects on trade. The first three of these capture the effects on, respectively, trade between members, member imports from nonmembers, and member exports to nonmembers. The other two are used to separate out the Japan-specific effects and capture an RTB's effects on member imports from Japan and on member exports to Japan. In equation (1), the RTB dummy variables are I^b_{ijt}, which is equal to one if i and j are both members of bloc b in year t; M^b_{ijt}, which is equal to one when j is a member of bloc b in year t (and i is not Japan); X^b_{ijt}, which is equal to one when i is a member of bloc b in year t (and j is not Japan); MJ^b_{ijt}, which is equal to one when j is a member of bloc b in year t and i is Japan; and XJ^b_{ijt}, which is equal to one when i is a member of bloc b in year t and j is Japan.

This specification of integration dummies differs from what is found in most of the gravity model literature in that, following Soloaga and Winters (2001), it includes dummies to capture the effects on member exports.[11] As reviewed by Soloaga and Winters, not only are member exports almost never considered, but many studies do not even consider the effects of RTBs on member imports, despite the prominence of trade diversion in any discussion of RTBs. Further, even when studies do consider member exports, they assume without any theoretical basis that the RTB will have the same effect on member exports as it will have on member imports.

As discussed above, the theoretical literature on the effects of RTBs on trade is rather sparse, so there are no definitive expected signs on the integration dummies in equation (1). Vinerian trade creation would suggest a positive sign for the effect on intra-bloc trade θ^b_1, and Vinerian trade diversion would suggest a negative sign for the effect on bloc-member exports out of the bloc θ^b_2 and θ^b_4. As mentioned above, several studies simply combine the effects on imports and exports, implying that θ^b_3 and θ^b_5 are also expected to have negative signs. However, in a standard general equilibrium model it is not possible to determine the signs of an RTB's effects on trade with nonmembers. As discussed above, further difficulty in assigning expected signs arises when you consider the geography of production decisions and changes in tariff levels following integration.

IV. THE EMPIRICAL RESULTS

To estimate equation (1), I use bilateral trade data from the *World Trade Flows* data set, GDP and population data from the World Bank, and inflation data from the Bureau of Labor Statistics. Using observations for four years (1982, 1987, 1992, and 1997) and 71 countries, I construct a balanced panel of 3,321 trading pairs per year for a total of 13,284 observations.[12]

In creating the integration dummies, some judgment must be made about the timing of the RTBs. This is because the date at which an RTB is actually implemented may not correspond to when it begins to affect trade. Trade may be affected even before an RTB is formally in place, as firms begin trading

[10] This is the least restrictive specification of fixed effects and follows Bayoumi and Eichengreen (1997) and Cheng and Wall (2001, 2002). In contrast, Mátyás (1997) assumes that each *country* has two fixed effects, one as an importer and one as an exporter. See Cheng and Wall (2002) for an evaluation of the different specifications.

[11] See Cheng and Wall (2001) for another multicountry study of the effects of RTBs that estimates separate effects for member imports and exports. In addition, Coughlin and Wall (2000) and Wall (2000) estimate the effects of NAFTA on member-country exports.

[12] Note that I do not have observations of trade between all pairs of these countries.

in advance of the barriers actually falling so as to establish themselves in new markets. Also, some RTBs, such as AFTA, are formalized well before trade barriers begin to fall significantly, while others lead to significant reductions in trade barriers immediately after formal implementation.

Keeping these problems in mind, my integration dummies are constructed according to the descriptions in Frankel (1997). An RTB is assumed to begin having an effect when it is formally implemented or when a country becomes a member of an existing bloc. Note that, because of evolving names and membership, the various North American and European regimes are collected into two meta-RTBs: the North American bloc and the European bloc.

Results for the least squares estimation of equation (1) are in Table 1, which provides the estimated coefficients, the White-corrected standard errors, and *t* statistics. In addition, the last column of the table provides the percentage change in trade implied by the point estimates of the coefficients on the integration and time dummies [$100 \times (e^{coeff.}-1)$]. Because of space considerations, I do not report the estimates of the 3,321 trading-pair fixed effects.

The estimated coefficients on the four gravity variables are not surprising and indicate that real trade was positively related with the countries' real GDPs and negatively related with their real per capita GDPs. Because of the perfect collinearity of the time dummies, the dummy for 1982 was excluded, meaning that the remaining time dummies indicated changes in trade relative to 1982.

As the results in Table 1 indicate, the six RTBs tended to have positive and statistically significant effects on their members' trade with each other. As for their effects on their members' trade with Japan and the rest of the world, there was not nearly as much consistency in sign or magnitude. Contrary to trade diversion, only one of the RTBs had a significant negative effect on imports from the rest of the world, although their effects on exports to the rest of the world did tend to be negative. As discussed in more detail below, the effects of the RTBs on members' trade with Japan tended to differ a great deal from their effects on members' trade with the rest of the world. As the present focus is Japan, the aggregated effects of the RTBs from Japan's perspective are provided in Table 2. These numbers are calculated by simply taking the point estimates of the effects of the RTBs on trade with Japan and applying them to the actual levels of trade for 1997.

A. North American Bloc

As reported in Table 1, I find that the North American trading bloc has had relatively large effects on all five categories of trade, especially intra-bloc trade, which was 57 percent higher because of the bloc. For trade with nonmembers, the bloc's effects were varied, affecting the members' trade with Japan differently from their trade with the rest of the world. Specifically, because of the bloc, member imports from the rest of the world were 18 percent higher, while member imports from Japan were 17 percent lower. Also, member exports to the rest of the world and to Japan were both lower because of the bloc, although the 37 percent decrease in exports to Japan was more than twice the effect on exports to the rest of the world.

As shown in Table 2, combining the negative effects that the North American bloc had on both directions of trade with Japan indicates that the bloc led to a decrease in total trade of US$53 billion in 1997, or 25 percent of members' total trade with Japan. Because the North American bloc is by far the most important RTB from Japan's perspective, this represented a significant decline in Japanese trade with the world. Specifically, because of the North American bloc, total Japanese exports and imports were, respectively, 5.4 percent and 9.2 percent lower, implying a 7 percent decrease in Japanese total trade with the world.

B. European Bloc

The estimated effects of the European bloc on trade are very different from those of the North American bloc. First, contrary to the predictions of Vinerian trade creation, I find that the European bloc had virtually no effect on trade between members. On the other hand, consistent with Vinerian trade diversion, I find that member imports from Japan and the rest of the world were, respectively, 36 percent and 11 percent lower because of the RTB. Interestingly, while member exports to the rest of the world were 22 percent lower, member exports to Japan were 31 percent higher.

As shown in Table 2, the opposing large effects of the European bloc on Japanese imports and exports meant that Japan's total trade with bloc members in 1997 was 8.7 percent, or US$10 billion, lower because of the bloc. In terms of Japan's trade with the world, the effect of the bloc was a 5.9 percent decrease in exports and a 4.6 percent increase in imports, for a decrease in total trade of 1.4 percent.

Table 1

Regression Results (Dependent Variable = Log of Real Exports)

	Coefficient	Robust s.e.	*t* statistic	Percentage change
Constant	–5.919	1.175	–5.038	
Log of real exporter GDP	1.244	0.208	5.985	
Log of real importer GDP	1.554	0.201	7.723	
Log of per capita exporter GDP	–0.780	0.192	–4.063	
Log of per capita importer GDP	–1.030	0.183	–5.644	
1987 dummy	–0.019	0.032	–0.584	–1.8
1992 dummy	–0.048	0.055	–0.868	–4.7
1997 dummy	0.045	0.075	0.603	4.6
North American bloc				
Intra-bloc trade	0.448	0.138	3.243	56.6*
Imports from rest of the world	0.167	0.060	2.785	18.2*
Exports to rest of the world	–0.188	0.055	–3.399	–17.2*
Imports from Japan	–0.196	0.092	–2.129	–17.8*
Exports to Japan	–0.454	0.213	–2.135	–36.5*
European bloc				
Intra-bloc trade	–0.025	0.063	–0.391	–2.4
Imports from rest of the world	–0.116	0.065	–1.779	–10.9
Exports to rest of the world	–0.248	0.044	–5.628	–21.9*
Imports from Japan	–0.444	0.103	–4.330	–35.9*
Exports to Japan	0.273	0.111	2.460	31.4*
AFTA				
Intra-bloc trade	0.322	0.135	2.387	38.0*
Imports from rest of the world	0.239	0.091	2.643	27.0*
Exports to rest of the world	0.307	0.075	4.102	35.9*
Imports from Japan	0.001	0.136	0.010	0.1
Exports to Japan	–0.096	0.166	–0.582	–9.2
ANZCERTA				
Intra-bloc trade	0.069	0.054	1.283	7.1
Imports from rest of the world	0.091	0.116	0.789	9.6
Exports to rest of the world	0.127	0.115	1.110	13.6
Imports from Japan	–0.684	0.113	–6.057	–49.5*
Exports to Japan	–0.388	0.085	–4.570	–32.1*
Mercosur				
Intra-bloc trade	0.334	0.108	3.101	39.6*
Imports from rest of the world	0.770	0.103	7.458	115.9*
Exports to rest of the world	–0.362	0.091	–3.966	–30.4*
Imports from Japan	–0.069	0.140	–0.490	–6.6
Exports to Japan	–0.381	0.214	–1.777	–31.7
Andean Community				
Intra-bloc trade	0.560	0.205	2.731	75.1*
Imports from rest of the world	0.230	0.083	2.769	25.9*
Exports to rest of the world	0.048	0.111	0.436	4.9
Imports from Japan	–0.520	0.214	–2.423	–40.5*
Exports to Japan	–0.238	0.354	–0.672	–21.1
Number of observations 13,284, $\bar{R}^2$= 0.898, RMSE = 0.907				

NOTE: The estimates of the 3,321 fixed effects are suppressed for space considerations. Asterisks indicate a percentage effect that is statistically significant at the 5 percent level.

Table 2

Aggregated Effects of Trade Blocs on Japanese Trade, 1997

	Actual 1997 trade (US$ billions)	Effect as percentage of trade with bloc	Effect in US$ billions	Effect as percentage of trade with world
North American bloc				
Japanese exports	130.6	–17.8	–23.2	–5.4
Japanese imports	80.9	–36.5	–29.5	–9.2
Total trade with Japan	211.5	–25.0	–52.8	–7.0
European bloc				
Japanese exports	69.9	–35.9	–25.1	–5.9
Japanese imports	47.5	31.4	14.9	4.6
Total trade with Japan	117.4	–8.7	–10.2	–1.4
AFTA				
Japanese exports	68.8	0.1	0.1	0.0
Japanese imports	48.7	–9.2	–4.5	–1.4
Total trade with Japan	117.5	–3.8	–4.4	–0.6
ANZCERTA				
Japanese exports	9.5	–49.5	–4.7	–1.1
Japanese imports	13.8	–32.1	–4.4	–1.4
Total trade with Japan	23.4	–39.2	–9.2	–1.2
Mercosur				
Japanese exports	4.8	–6.6	–0.3	–0.1
Japanese imports	4.0	–31.7	–1.3	–0.4
Total trade with Japan	8.7	–18.0	–1.6	–0.2
Andean Community				
Japanese exports	2.4	–40.5	–1.0	–0.2
Japanese imports	1.2	–21.1	–0.2	–0.1
Total trade with Japan	3.6	–34.1	–1.2	–0.2
Total for all blocs				
Japanese exports	286.0	–19.0	–54.3	–12.7
Japanese imports	196.0	–12.8	–25.0	–7.8
Total trade with Japan	482.1	–16.5	–79.3	–10.6

NOTE: The trade data in the second column are from *World Trade Flows,* and the numbers in the shaded areas are from the last column of Table 1. Other numbers are the author's calculations.

C. AFTA

I find that because of the formation of AFTA, intra-bloc trade and members' trade with the non-Japanese world were both much higher. Intra-bloc trade increased by 38 percent, while imports from and exports to the rest of the world rose by 27 percent and 36 percent, respectively. On the other hand, the estimated effects of AFTA on trade with Japan were both statistically no different from zero, leaving Japan out of the AFTA trade boom.

In terms of trade volume, AFTA members are collectively as important to Japan as the members of the European bloc. Nevertheless, by 1997, AFTA did not have much of an effect on Japan. Clearly, though, the jury is still out on the effects of AFTA on Japanese trade as the relative newness and shallowness of AFTA integration are likely responsible

for both the small estimated effects and the large standard errors. From a Japanese perspective, this RTB is perhaps the most interesting for future research, because the effects of the other two large blocs are probably already firmly established.

D. ANZCERTA

The only significant effects that I find for the ANZCERTA are the large decreases in member trade with Japan. As with AFTA and NAFTA, this agreement has been much worse for Japan than for the rest of the world, having relatively little effect on trade between members and the rest of the world other than Japan. It has decreased members' imports from Japan by 50 percent and member exports to Japan by 32 percent. In total, this indicated a 39 percent drop in members' total trade with Japan and a 1.2 percent decrease in Japan's total trade with the world.

E. Mercosur

The estimated effects of Mercosur are fairly large and pronounced, even though this RTB was formed relatively recently. I find that, because of Mercosur, trade between members was 40 percent higher, imports from the rest of the world were 116 percent higher, and exports to the rest of the world were 30 percent lower. It had very little effect on its members' imports from Japan, although it decreased their exports to Japan by 32 percent. Combining these effects, members' total trade with Japan was 18 percent lower because of Mercosur. Since members of this RTB accounted for less than 2 percent of Japanese trade with the world, it has not had a large effect on Japan.

F. Andean Community

As with most of the other RTBs, the formation of the Andean Community has led to a large increase in trade between members and imports from the rest of the world. Japan has not shared in this, however, as member imports from Japan were 41 percent lower in 1997 because of the RTB. Because members of the Andean Community's share of Japanese exports was less than 1 percent of the total, this RTB has so far had little effect on Japan's overall trade.

G. Total for All Blocs

To illustrate how regional integration as a whole has affected Japanese trade, the bottom of Table 2 presents the aggregated effects of the six RTBs on 1997 trade. Primarily because of the large negative effects of the two largest RTBs, the North American and European blocs, the total effect on Japanese exports to RTB members was a decrease of 19 percent, or US$54 billion. On the imports side, the large positive effect of the European bloc softened the large negative effect of the North American bloc, making the total effect in Japanese imports from RTB members a 12.8 percent, or US$25 billion, decrease. The combined decreases in Japanese exports and imports meant a 16.5 percent, or US$79 billion, decrease in Japanese total trade with members of these RTBs.

Because the members of these six RTBs accounted for more than 60 percent of Japan's trade, the total effects of the RTBs were significant in terms of their effects on total Japanese trade with the world. Specifically, their total effects meant a 12.7 percent decrease in Japanese exports, a 7.8 percent decrease in Japanese imports, and a 10.6 percent decrease in Japanese total trade.

V. CONCLUDING REMARKS

This study is a first step in understanding the effects of regional integration on Japan. Despite recent trends toward regional integration, Japan, the world's second largest economy, has so far resisted joining an RTB. Because more than 60 percent of Japan's trade is with countries that are members of a major RTB, its reluctance may have had significant effects on its pattern and volume of trade. Indeed, I find that Japanese trade, especially Japanese exports, has been reduced by the regional integration of its trading partners. Specifically, I find that if none of these RTBs were in place, Japan's 1997 total trade volume would have been nearly 11 percent higher than it was—exports and imports would have been almost 13 and 8 percent higher, respectively. As noted above, there is still some uncertainty about the eventual effects of some of the RTBs, as several are relatively new and not nearly as deeply integrated as the North American and European blocs. In particular, as the members of AFTA become more integrated over time, its effects on Japan are likely to become clearer and more pronounced than those found in this study.

Perhaps the most curious aspect of my results, though, is that the effects of the RTBs on Japan tended to differ a great deal from their effects on the rest of the world, almost always being much more negative. This finding presents the most obvious direction for future research, which should include sectoral- or industry-level analysis, with

attention paid to the differences between intermediate and final goods. This is likely to be particularly important given the high mobility of Japanese firms and capital and the endogeneity of the production-location decision. Also, although Head and Ries (2001) find that Japanese exports of intermediate goods tend to follow manufacturing investment overseas, for key firms this was not true. Overseas investment by large automobile assemblers such as Toyota, Nissan, and Honda instead led to net decreases in their exports. Similarly, Lipsey, Ramstetter, and Blomström (2000) find differences across Japanese industries.

REFERENCES

Abe, Kenzo and Zhao, Laixun. "International Joint Ventures, Economic Integration, and Government Policy," in Takatoshi Ito and Anne O. Krueger, eds., *The Role of Foreign Direct Investment in East Asian Economic Development*. Chicago: University of Chicago Press, 2000.

Anderson, Kym and Snape, Richard H. "European and American Regionalism: Effects on and Options for Asia." *Journal of the Japanese and International Economies*, December 1994, *8*(4), pp. 454-77.

Bandyopadhyay, Subhaya and Wall, Howard J. "Customs Union or Free Trade Area? The Role of Political Asymmetries." *Review of International Economics*, November 1999, *7*(4), pp. 665-72.

Bayoumi, Tamim and Eichengreen, Barry. "Is Regionalism Simply a Diversion? Evidence from the Evolution of the EC and EFTA," in Takatoshi Ito and Anne O. Krueger, eds., *Regionalism versus Multilateral Trade Arrangements*. Chicago: University of Chicago Press, 1997.

Ceglowski, Janet. "The Recent Behavior of Japanese Imports: A Disaggregated Analysis." *Japan and the World Economy*, December 1996, *8*(4), pp. 443-57.

Cheng, I-Hui and Wall, Howard J. "Estimates of the Effects of Regional Integration on Trade." Working paper, Federal Reserve Bank of St. Louis, 2001.

_____ and _____. "Controlling for Heterogeneity in Gravity Models of Trade and Integration." Working Paper 99-010C, Federal Reserve Bank of St. Louis, 2002.

Coughlin, Cletus. C. and Wall, Howard J. "NAFTA and the Changing Pattern of State Exports." Working Paper 2000-029A, Federal Reserve Bank of St. Louis, February 2000 (forthcoming in *Papers in Regional Science*).

Daly, Kevin. "Does Exchange Rate Volatility Impede the Volume of Japan's Bilateral Trade?" *Japan and the World Economy*, July 1998, *10*(3), pp. 333-48.

Deardorff, Alan V. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in Jeffrey A. Frankel, ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998.

Eaton, Jonathan and Tamura, Akiko. "Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns," *Journal of the Japanese and International Economies*, December 1994, *8*(4), pp. 478-510.

_____ and _____. "Japanese and U.S. Exports and Investment as Conduits of Growth," in Anne O. Krueger and T. Ito, eds., *Financial Deregulation and Integration in East Asia*. Chicago: University of Chicago Press, 1996.

Eguchi, Masato. "JSEPA (Japan-Singapore Economic Partnership Agreement) and Future Trade Policy." *Journal of Japanese Trade and Industry*, January/February 2001, *20*(1).

Feenstra, Robert C. "World Trade Flows, 1980-1997." Center for International Data, University of California, Davis, March 2000.

Frankel, Jeffrey A. *Regional Trading Blocs in the World Economic System*. Washington, DC: Institute for International Economics, 1997.

Goto, Junichi and Hamada, Koichi. "Economic Integration and the Welfare of Those Who Are Left Behind: An Incentive Theoretic Approach." *Journal of the Japanese and International Economies*, 1998, *12*(1), pp. 25-48.

_____ and _____. "Regional Economic Integration and Article XXIV of the GATT." *Review of International Economics*, November 1999, *7*(4), pp. 555-70.

Greaney, Theresa. "An Analysis of Japan's Changing Import Behavior." Working paper, University of Hawaii, Manoa, 2001.

Head, Keith and Ries, John. "Overseas Investment and Firm Exports." *Review of International Economics*, February 2001, *9*(1), pp. 108-22.

Kimura, Fukanari. "Location and Internalization Decisions: Sector Switching in Japanese Outward Foreign Direct

Investment," in Takatoshi Ito and Anne O. Krueger, eds., *The Role of Foreign Direct Investment in East Asian Economic Development*. Chicago: University of Chicago Press, 2000.

Kojima, Akira. "Free Trade Agreements as Constructive Regionalism." *Journal of Japanese Trade and Industry*, January/February 2001, *20*(1).

Kose, M. Ayhan and Riezman, Raymond. "Understanding the Welfare Implications of Preferential Trade Agreements." *Review of International Economics*, November 2000, *8*(4), pp. 619-33.

Linnemann, Hans. *An Econometric Study of International Trade Flows*. Amsterdam: North-Holland, 1966.

Lipsey, Robert E. "Affiliates of U.S. and Japanese Multinationals in East Asian Production and Trade," in Takatoshi Ito and Anne O. Krueger, eds., *The Role of Foreign Direct Investment in East Asian Economic Development*. Chicago: University of Chicago Press, 2000.

_____; Ramstetter, Eric and Blomström, Magnus. "Outward FDI and Parent Exports and Employment: Japan, the United States, and Sweden." *Global Economy Quarterly*, September-October 2000, *1*, 285-302.

Mátyás, Laszlo. "Proper Econometric Specification of the Gravity Model." *The World Economy*, May 1997, *20*(3), pp. 363-68.

Ministry of Economy, Trade, and Industry. *2001 Report on the WTO Consistency of Trade Policies by Major Trading Partners*. 2001.

Ministry of International Trade and Industry. *White Paper on International Trade, 1998*. 1998.

Nadenichek, Jon. "The Japan-U.S. Trade Imbalance: A Real Business Cycle Perspective." *Japan and the World Economy*, August 2000, *12*(3), pp. 255-71.

Oguledo, Victor I. and MacPhee, Craig R. "Gravity Models: A Reformulation and an Application to Discriminatory Trade Arrangements." *Applied Economics*, February 1994, *26*(2), pp. 107-20.

Pöyhönen, P. "A Tentative Model for the Volume of Trade Between Countries." *Weltwirtschaftliches Archive*, 1963, *90*, pp. 93-100.

Sawyer, W. Charles and Sprinkle, Richard L. "The Demand for Imports and Exports in Japan: A Survey." *Journal of the Japanese and International Economies*, June 1997, *11*(2), pp. 247-59.

Soloaga, Isidro and Winters, L. Alan. "Regionalism in the Nineties: What Effect on Trade?" *The North American Journal of Economics and Finance*, March 2001, *12*(1), pp. 1-29.

Tinbergen, Jan. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund, 1962.

Wall, Howard J. "NAFTA and the Geography of North American Trade." Working Paper 2000-17B, Federal Reserve Bank of St. Louis, November 2000.

Winters, L. Alan. "Regionalism and the Rest of the World: The Irrelevance of the Kemp-Wan Theorem." *Oxford Economic Papers*, April 1997, *49*(2), pp. 228-34.

_____ and Chang, Won. "Regional Integration and Import Prices: An Empirical Investigation." *Journal of International Economics*, August 2000, *51*(2), pp. 363-77.

Viner, Jacob. *The Customs Union Issue*. Carnegie Endowment for International Peace, 1950.

Yi, Sang-Seung. "Free Trade Areas and Welfare: An Equilibrium Analysis." *Review of International Economics*, May 2000, *8*(2), pp. 336-47.

The FOMC's Balance-of-Risks Statement and Market Expectations of Policy Actions

Robert H. Rasche and Daniel L. Thornton

In January 2000 the Federal Open Market Committee (FOMC) instituted the practice of issuing a "balance of risks" statement along with its policy decision at the close of each FOMC meeting. The balance-of-risks statement was intended to indicate the Committee's assessment of the balance of risks for heightened inflation pressures or economic weakness over the foreseeable future. In announcing the procedural change, the FOMC explicitly noted that "this time frame in the new language is intended to cover an interval extending beyond the next FOMC meeting," suggesting that the balance-of-risks statement should not be interpreted as an indicator of the Committee's next policy action.

Previously, the FOMC had included a statement in its policy directive that appeared to pertain to possible future policy actions and came to be known as the "symmetry," "tilt," or "bias." The directive was said to be symmetric, or unbiased, if the directive indicated that a tightening or easing of policy was equally likely during the period between FOMC meetings—the "intermeeting period."

The purpose of this article is to review the FOMC's use of its balance-of-risks statement and the market's interpretation of it. Despite the FOMC's claim that the balance-of-risks statement is not intended to signal any particular action at or before the next FOMC meeting, market participants have used the statement when assessing the likelihood of a policy action at the next meeting.

HISTORICAL DEVELOPMENT OF FOMC STATEMENTS

The practice of adopting a bias in the policy directive began in 1983; however, until 1999, the statement of the bias that the FOMC adopted at one meeting was not made public until after the next meeting. In May 1999, the FOMC changed its practice and began announcing the symmetry of its policy directive at the conclusion of each meeting. This announcement attracted considerable attention. While the FOMC had never offered a formal interpretation of the symmetry clause of its policy directive, the market interpreted the bias in one direction or another as an indication of the likelihood that the FOMC would change the intended funds rate in that direction.[1] This interpretation was reasonable given that the bias was stated in terms of the need to change the degree of pressure in reserve markets, i.e., the intended funds rate, during the intermeeting period. Nevertheless, Thornton and Wheelock (2000) found that the bias had essentially no predictive content for changes in the funds rate target at or before the next meeting. Specifically, they found that, while any action taken was nearly always in the direction of the bias at the previous meeting, they could not reject the hypothesis that policy actions taken were independent of the asymmetric language adopted at the previous meeting. Consequently, the evidence suggests that policymakers were no more likely to change the intended funds rate when the bias at the previous meeting was asymmetric.

Nevertheless, it appears that the FOMC was concerned that immediate release of the bias was giving rise to undue expectations of a policy action at or before the next FOMC meeting. Consequently, in announcing its new procedure on January 19, 2000, the FOMC emphasized that the balance-of-risks statement was not intended to convey information about future policy actions. Specifically, the FOMC noted that, "previously, the Committee's directive and statement referred to the relative likelihood of an increase or a decrease in the intended federal funds rate, which may have intensified the public focus on the chance of a subsequent adjustment to the stance of policy, thereby increasing the possibility of misperceptions about the odds and timing of policy action."[2]

By removing explicit reference to both the intended federal funds rate and the intermeeting period, the Committee hoped that the new language would not be interpreted as indication of the likelihood of a policy action at or before the next scheduled Committee meeting.

Robert H. Rasche is a senior vice president and director of research and Daniel L. Thornton is a vice president and economic advisor at the Federal Reserve Bank of St. Louis. Charles Hokayem and Julie Knoll provided research assistance.

[1] For a discussion of three possible interpretations of the "tilt," see Thornton and Wheelock (2000).

[2] FOMC (January 19, 2000).

POLICY ACTIONS AND THE FOMC'S BALANCE-OF-RISKS STATEMENT

The FOMC increased its target for the funds rate three times in 2000. All of these changes occurred at regularly scheduled FOMC meetings (February, March, and May), and all were made when the balance-of-risks statement adopted at the preceding meeting indicated heightened inflation pressures.

The Committee's target for the federal funds rate was reduced eleven times during 2001, once at each of the eight regularly scheduled FOMC meetings and three times between scheduled meetings. Again, on each of these occasions the Committee had indicated at the preceding meeting that the risks were toward economic weakness.

Consequently, all 14 of the target changes that occurred in 2000 and 2001 were made after the Committee had indicated that the balance of risks were weighted in the direction consistent with the next target rate change.

The FOMC and the Balance-of-Risks Statement

To assess the market's interpretation of the balance-of-risks statement, we read press analyses of the FOMC's public statement following each meeting and other analyses of monetary policy over the period since the adoption of the new procedure. The relevant sources and quotations are presented in the appendix. It is clear from reading these accounts that the balance-of-risks statement was one of the pieces of information that market participants used to determine the likelihood of an action at the next meeting.[3] For example, on May 17, 2000—the day after the FOMC announced that it was increasing its target for the intended federal funds rate by 50 basis points, the largest change in the funds rate target in over five years—the *Los Angeles Times* reported that the FOMC "hinted that it may do so again next month," noting that the Fed's hint of further rate increases came "in the form of a warning that inflation remains a serious risk."[4] Similar statements appeared in several other major newspapers on that day.

This interpretation is not unexpected since the "foreseeable future" language of the balance-of-risks statement includes the period up to and including the next regularly scheduled FOMC meeting. This interpretation was likely exacerbated by the FOMC's use of the balance-of-risks statement, as well as statements made by some members of the FOMC.

An example of this occurred in late 2000 and early 2001, when the FOMC changed its balance-of-risks statement for the first time and soon after made an intermeeting move: For the first seven FOMC meetings in 2000, the Committee indicated that the balance of risks were "mainly toward conditions that may generate heightened inflation pressures in the foreseeable future." At the eighth meeting, on December 19, 2000, the FOMC reversed the balance-of-risks statement, indicating that "it believes that the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future." Despite this dramatic swing from unbalanced risks in one direction to unbalanced risks in the other, the Committee chose not to change the intended federal funds rate at that meeting, leaving it at 6.5 percent. About two weeks later, on January 3, 2001, the FOMC reduced the funds rate objective by 50 basis points in an intermeeting move.

Whether intended or not, these actions may have conditioned market participants to believe that the balance-of-risks statement was a good indicator of the FOMC's next policy action. Indeed, the minutes of the December 19, 2000, FOMC meeting indicate that some Committee members thought that the shift in the balance-of-risks statement would have this effect. The minutes note that "the revised statement of risks, even though it would not be associated with an easing move, could strengthen expectations regarding future monetary policy easing to an extent that was difficult to predict and could generate sizable reactions in financial markets."[5]

In any event, press reports show that the 180-degree swing in the balance-of-risks statement was widely interpreted as a signal that the Fed would reduce rates at the next FOMC meeting in late January. More than a year later, in reporting on the events surrounding the December 19, 2000, FOMC meeting, the *Financial Times* noted that at least one member was thinking of the action as a signaling device for policy actions, noting that, "as one FOMC member says, policy has become as much about

[3] In 1994 the FOMC began the practice of adjusting its funds rate target primarily at regularly scheduled FOMC meetings. Consequently, most expectations were for changes at a regularly scheduled meeting and not during the intermeeting period. Indeed, the evidence here and in Poole, Rasche, and Thornton (2002) shows that the markets were surprised by intermeeting moves.

[4] Mulligan (2000).

[5] FOMC (2001, p. 229).

signaling future rate changes as about actual immediate rate changes." The article goes on to say that "Mr. Moskow, of the Chicago Fed, a keen bridge player, describes the change in tilt as a 'jump shift,' a signal of a powerful hand."[6]

The idea that the Committee viewed the shift in the balance-of-risks statement as a signal of likely future policy actions is borne out in other ways. For example, there was a little-noticed sentence in the December 19, 2000, press release, stating that "the Committee will continue to monitor closely the evolving economic situation." The *Financial Times* later noted that this "was a piece of classic Fedspeak—an apparently anodyne and rather obvious observation that the Fed was on heightened alert, and would not necessarily wait six weeks until the next scheduled meeting to cut interest rates." The article goes on to quote President Santomero of the Federal Reserve Bank of Philadelphia: "We indicated that the situation was sufficiently fluid that we were paying special attention to new information that was coming up on a week to week basis."

That the FOMC was poised to move ahead of the January 30/31, 2001, meeting is also reflected in the minutes of the December 19, 2000, meeting which stated that, "on balance, the information already in hand indicated that the expansion clearly was weakening and by more than had been anticipated. In the circumstances, prompt and forceful policy action sooner and larger than expected by financial markets seemed called for."

The view that the Fed would eventually have to reduce the funds rate target was widely held. Indeed, by mid-December 2000, the federal funds futures market was already pricing-in a significant probability of a 50-basis-point decline in the funds rate in late January 2001; within a few days of the December 2000 meeting, the February 2001 federal funds rate futures contract was essentially fully pricing-in a 50-basis-point reduction in the funds rate at the January 30/31, 2001, meeting. The FOMC surprised the market by reducing the funds rate by 50 basis points on January 3, 2001. Moreover, it again adopted a balance-of-risks statement indicating the prospects for economic weakness. The futures market almost immediately priced in another 50-basis-point cut at the Fed's regularly scheduled January 30/31 meeting, dropping 29 basis points on January 3 and another 19 basis points on January 4.[7]

Further evidence that some Committee members interpreted the balance-of-risks statement as a signal of future policy actions came with the release of the minutes of the May 15, 2001, FOMC meeting. The minutes of that meeting report that "the members anticipated that a neutral balance of risks statement could be appropriate before long, probably well before substantial evidence had emerged that economic growth had strengthened appreciably, once the Committee could see that policy had eased enough to promote a future return to maximum sustainable economic growth." In reporting on these minutes, John Berry of the *Washington Post* noted that the end of the easing process "would be marked by a statement from the committee that it had decided that a 'neutral balance of risks' had been achieved."[8]

It is clear from press reports that during 2001 the balance-of-risks statement was an important indicator of a likely Fed action. It is equally clear that this interpretation was intensified by statements of some members of the FOMC and by the FOMC's use of the balance-of-risks statement. A recent example can be seen by contrasting the remarks Chairman Greenspan made in a speech in San Francisco on January 11, 2002, with his testimony on the state of the economy given to the Senate Budget Committee on January 24, 2002. In the San Francisco speech, using phraseology similar to the balance-of-risks statement, the Chairman said, "I would emphasize that we continue to face significant risks in the near term. Profits and investment remain weak and, as I noted, household spending is subject to restraint from the backup in interest rates, possible increases in unemployment, and from the effects of widespread equity asset price deflation over the past two years."[9] His testimony before the Senate Budget Committee was more upbeat. On January 25, 2002, the *Wall Street Journal* reported the following: "In a rare admission of miscalculation for a man considered the master market manipulator, the Fed chairman told Congress that…by making a statement in mid-January like 'we continue to face significant risks in the near term,' Mr. Greenspan later realized that he had unintentionally 'implied that I didn't think the economy was in the process of turning.'"[10] The *Wall Street Journal* noted further: "Just as Mr.

[6] Baker (2001).

[7] See Poole, Rasche, and Thornton (2002, appendix) for details.

[8] Berry (2001).

[9] Greenspan (2002).

[10] Schlesinger (2002).

Table 1

FOMC Events and a Measure of Unexpected Policy Action

Date	Poole/Rasche Measure
2/02/00	–0.04
3/21/00	–0.01
5/16/00	0.04
6/28/00*	–0.02
8/22/00*	0.00
10/03/00*	0.00
11/15/00*	0.00
12/19/00*	0.05
1/03/01†	–0.29
1/31/01	0.00
3/20/01	0.03
4/18/01†	–0.42
5/15/01	–0.07
6/28/01	0.02
8/21/01	0.02
9/17/01†	–0.20
10/02/1	–0.08
11/6/01	–0.11
12/11/01	0.01
1/30/02*	0.02

NOTE: *Meeting, no change in the intended funds rate.
†Indicates an intermeeting target change.

Greenspan's mid-January speech led many analysts to expect one more Fed interest rate cut later this month," based on a more balanced assessment of the economic outlook, "yesterday's remarks persuaded many Fed watchers to revise their forecast for the Jan. 29-30 monetary policy meeting."[11] Hence, despite the fact that the risks were weighted toward economic weakness, Greenspan's comments appeared to persuade market participants that no additional easing was likely in late January.

THE IMPORTANCE OF THE BALANCE-OF-RISKS STATEMENT FOR MARKET EXPECTATIONS

While there is little doubt that the market considers the balance-of-risk statement in determining the likelihood of the next policy action, the important question is how important is this information for determining market expectations of a policy move. One way is to see whether the balance-of-risks statement helps the market to correctly anticipate the FOMC's actions. To do this, we use the measure from Poole and Rasche (2000) of unexpected changes in the intended federal funds rate. The Poole/Rasche measure uses the change in the 1-month-ahead federal funds futures rate on the day the target was changed as their measure of the unexpected change in the intended funds rate. For the first day of the month, they use the difference between the rate on the 1-month futures rate on the first day of the month and the rate on the 2-month futures contract on the last day of the previous month. (See Poole and Rasche, 2000, and Poole, Rasche, and Thornton, 2002, for details.)

The Poole/Rasche measure of the unexpected change is reported in Table 1 for each meeting and for the three intermeeting changes during our sample period. Because of ambient variation in the federal funds futures rate, changes of 5 basis points or less are considered insignificant. The Poole/Rasche measure suggests that there were no instances during 2000 when there was an unexpected action by the FOMC. During the first three meetings, the intended funds rate was raised when the balance of risks were weighted toward heightened inflation pressures. In each of these cases, the market appears to have anticipated the FOMC's action.

The results for 2001 prior to September 11 were very similar. Market participants were only surprised when the FOMC made intermeeting changes in the intended funds rate. On all of these occasions, however, market participants anticipated that the FOMC would reduce the intended funds rate; they were surprised only by the timing of the action. It appears to be difficult to determine the precise day when the FOMC will take an intermeeting action, even if the market believes that such an action is likely.

There was a 7-basis-point change in the federal funds futures rate on May 15, when the FOMC reduced the intended funds rate by 50 basis points. Poole, Rasche, and Thornton (2002) note, however, that a more detailed analysis of news reports and the futures rate shows that this change was in fact anticipated. The other large change in the futures rate occurred in the two months following the terrorist attacks on September 11. Figures 1 and 2 present the daily rates on the November and December

[11] Schlesinger (2002).

Figure 1

November 2001 Federal Funds Futures Rate

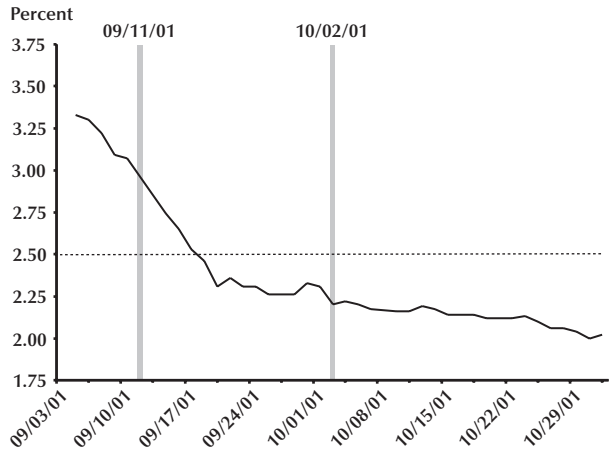
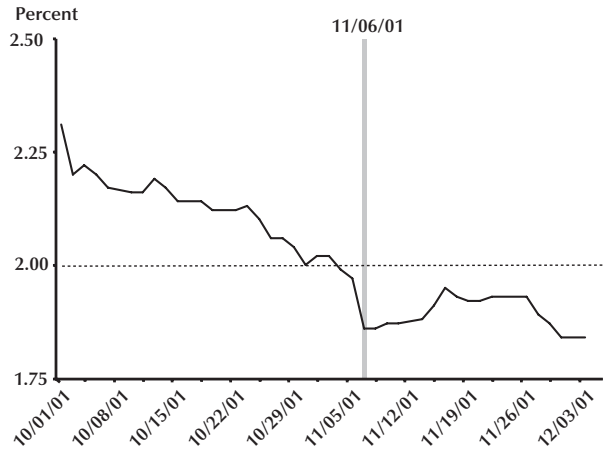


Figure 2

December 2001 Federal Funds Futures Rate



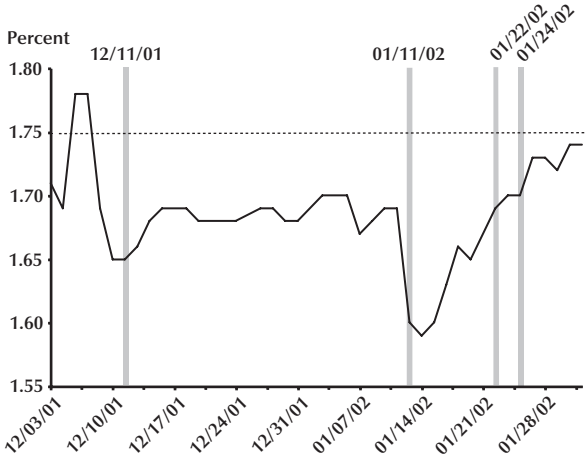
federal funds futures contracts around the dates of the October and November FOMC meetings, respectively. Vertical lines indicate the dates of these meetings and the terrorist attack. Consistent with news reports, these figures suggest that the cuts in the intended federal funds rate were effectively anticipated by the time of the meeting. The unusually large changes in the futures rates associated with these meetings may be due in part to the greater uncertainty in the wake of the terrorist attacks. This uncertainty is particularly evident in the December contract.

Was the balance-of-risks statement definitive for correctly forecasting policy actions? The answer is, apparently not. Table 1 shows that the market correctly anticipated that the FOMC would not change the intended funds rate at each of the last five FOMC meetings during 2000 despite the fact that, on each of these occasions, the balance-of-risks statements were also weighted toward heightened inflation pressures. Hence, while the press analyses suggest that market participants look to the balance-of-risk statement as one source of information, it is not the only source. Indeed, it appears that it may not be a critical source of information.

Market participants apparently also rely on their understanding of how the FOMC will respond to the latest economic reports and on statements of the Chairman and other Fed officials. Perhaps the clearest example of the latter is shown by the behavior of the rate on the February 2002 federal funds futures contract. This contract is used because the January

Figure 3

February 2002 Federal Funds Futures Rate



2002 meeting was scheduled for January 29-30, so the market's expectations of FOMC actions at this meeting are best reflected in the February contract.

The daily rate on this contract is plotted in Figure 3. The four vertical lines denote the dates of the December 2001 FOMC meeting, Chairman Greenspan's San Francisco speech, the Tuesday following two press reports of a statement by a senior Federal Reserve official clarifying the Chairman's San Francisco remarks, and the Chairman's Senate testimony. The behavior of the federal funds futures rate suggests that there was considerable uncertainty about the February 2002 federal funds rate before

the December 11, 2001, FOMC meeting. After that meeting, when the FOMC indicated that the risks remained balanced toward economic weakness, expectations for the funds rate settled down and the market priced-in a significant probability of a 25-basis-point reduction in the intended funds rate at the January 2002 meeting.

Market expectations were significantly affected by Chairman Greenspan's remarks in San Francisco on January 11, 2002, when the February federal funds futures rate fell below 1.6 percent. Market expectations were quickly revised, reducing the probability of an additional cut in the intended funds rate, and were revised further on Tuesday, January 22, 2002, in the wake of two reports—one in the *Washington Post* on Saturday, January 19, and the other in the *Wall Street Journal* on Monday, January 21—quoting an "unidentified Fed official" as saying that the Chairman's downbeat remarks in San Francisco were overplayed and that a rate cut was not all that likely. The probability of a cut in the intended funds rate was all but eliminated following the Chairman's testimony before the Senate Budget Committee on January 24, 2002.

Interpreting the Balance-of-Risks Statement

While market participants appear to rely on a wide range of information to determine the likelihood of a policy action, it is interesting to note that, on all 14 occasions when the funds rate was adjusted, the balance-of-risks statement was unbalanced in the direction of the rate change. Moreover, in all but one of the cases where the balance-of-risks statement was unbalanced toward economic weakness, the FOMC decreased the policy rate. The exception occurred in January 2002, when the FOMC said that the balance of risks was unbalanced toward weakness, but took no action. In its public statement, however, the FOMC noted that "signs that weakness in demand is abating and economic activity is beginning to firm have become more prevalent. With the forces restraining the economy starting to diminish, and with the long-term prospects for productivity growth remaining favorable and monetary policy accommodative, the outlook for economic recovery has become more promising." Hence, despite the fact that risks were slanted toward economic weakness, the FOMC made it clear that indications were that the economy was strengthening.

We speculate that the fact that the FOMC has nearly always reduced the funds rate objective when the balance-of-risk statement was unbalanced toward economic weakness is a natural consequence of the statement itself. It is difficult to see why the Committee would not act promptly in an attempt to offset these risks to whatever extent possible when policymakers believe that the risks are tilted toward economic weakness. Indeed, in reacting to the January 3, 2001, funds rate cut, former Governor Wayne Angell echoed this sentiment, saying "I've never seen the Fed get themselves into such a dilemma as they were in Dec. 19, saying how bad the economy was but also saying they weren't acting [then]. After that, they needed to cut rates, probably within the first two weeks [of the year]."[12] In retrospect, we now know that, properly interpreted, the remainder of the statement made it clear that the FOMC had no intention of waiting very long to reduce the funds rate target.

One could argue that the same argument applies to situations where the balance of risks is weighted toward heightened inflation pressures. We believe there are differences, however. For one thing, signs of slowing in economic growth are typically readily apparent—a rising unemployment rate, reductions in production and/or sales, weakening consumer and investor confidence, etc. Indicating that these signs of a weakening economy are emerging and threaten to worsen will naturally lead the public to expect that policymakers will take actions to prevent a downturn.

In contrast, signs that inflation may be worsening are more amorphous. Forward-looking inflation indicators—the spread between inflation adjusted and non-inflation adjusted Treasury rates, money growth measures, inflation surveys, and commodities futures prices—have not proven reliable predictors of near-term inflation. Policymakers would be more likely to act if the underlying trend in inflation were rising; however, month-to-month inflation numbers are quite volatile. It generally takes several months to obtain conclusive evidence of a significant shift in the underlying inflation rate. Consequently, market participants might be less inclined to believe that the FOMC will react quickly to a "bad" inflation report.

Moreover, former Vice Chairman Blinder (1998, pp. 19-20) has suggested that a central bank "will take far more political heat when it tightens preemptively to avoid higher inflation than when it eases preemp-

[12] Schlesinger, Ip, and Kulish (2001); see appendix.

tively to avoid higher unemployment." If true, then the "political" costs of adjustment are higher for raising the funds rate target than for lowering it. Such asymmetric adjustment costs may generate more inertia in "tightening" than in "easing" policy.

For these reasons, we believe that market participants are less likely to interpret a statement that the risks are weighted toward heightened inflation pressures as an indicator of an impending FOMC tightening than to interpret a statement that the risks are weighted toward economic weakness as an indicator of an impending FOMC easing.

CONCLUSION

The balance-of-risks statement is only one of the factors that market participants consider in forming their expectations of FOMC actions, and it appears that this statement alone is not a critical factor. An important source of information is the "clarifying statement" that sometimes accompanies the announcements made at the conclusion of FOMC meetings, as well as general statements made by the Chairman and other FOMC members. The importance of the accompanying statement was apparently not appreciated initially. It appears that the FOMC's attempt to signal that it would likely take action before its regularly scheduled FOMC meeting on January 31, 2001, was too cryptic. Later, the importance of the statement was recognized. A similar message sent in the statement following the January 2002 meeting was not misinterpreted. In spite of the statement that the risks remained unbalanced toward economic weakness, this FOMC statement was widely interpreted to mean that no additional easing actions were likely to occur in the absence of significant new evidence.

We believe that at each meeting the FOMC policymakers set the policy instrument at the level that they believe to be consistent with their policy objectives given what they then know about the state of the economy. Policymakers should be prepared to act when new information suggests that their economic objectives cannot be obtained without adjusting the policy instrument. Policymakers might do well to indicate the kinds of information they believe to be important in making these decisions. Over time, the combination of (i) information about what policymakers believe to be important and (ii) their reaction to economic reports will provide market participants a better framework for anticipating policy actions—an activity that is certain to continue regardless of the FOMC's disclosure policy.

REFERENCES

Baker, Gerard. "The Fed's Extraordinary Year: The Fed's Fight to Stay Ahead of the Game." *Financial Times*, 18 December 2001. <FT.com>.

Berry, John M. "Fed Minutes Hint Cuts Are Nearly Over." *Washington Post*, 12 June 2001, p. E01.

Blinder, Alan S. *Central Banking in Theory and Practice*. Cambridge, MA: MIT Press, 1998.

Federal Open Market Committee. "Modification of the FOMC's Disclosure Procedures." Board of Governors of the Federal Reserve, 19 January 2000.

_____. "Minutes of the Federal Open Market Committee, December 19, 2000, *Federal Reserve Bulletin*, April 2001, pp. 224-30.

Greenspan, Alan. "*The Economy*." Presented at the Bay Area Council, San Francisco, California, 11 January 2001. <www.federalreserve.gov/boarddocs/speeches/2002>.

Mulligan, Thomas S. "Fed Targets Inflation, Hikes Rate Half-Point; Economy: Central bank hints at still more tightening. Banks quickly raise their prime rates, meaning consumers will soon feel the pinch." *Los Angeles Times*, 17 May 2000, p. A-1.

Poole, William and Rasche, Robert H. "Perfecting the Market's Knowledge of Monetary Policy." *Journal of Financial Services Research*, December 2000, *18*(2-3), pp. 225-98.

_____; _____ and Thornton, Daniel L. "Market Anticipations of Monetary Policy Actions." Federal Reserve Bank of St. Louis *Review*, July/August 2002, *84*(4), pp. 65-94.

Schlesinger, Jacob M.; Ip, Greg and Kulish, Nicholas. "Two-Edged Sword: Fed's Surprise Move Sparks Market Rally, Sets Off New Jitters." *Wall Street Journal*, 4 January 2001, p. A1.

_____. "Greenspan Offers a Brighter U.S. Forecast." *Wall Street Journal*, 25 January 2002, p. A2.

Thornton, Daniel L. and Wheelock, David C. "A History of the Asymmetric Policy Directive." Federal Reserve Bank of St. Louis *Review*, September/October 2000, *82*(5), pp. 1-16.

Appendix

PRESS ANALYSES OF THE FOMC'S PUBLIC STATEMENTS FOLLOWING EACH FOMC MEETING: JANUARY 2000 THROUGH JANUARY 2002

02/03/2000
Stocks Mixed on Rate News; Treasury's Plan Rallies Bonds; Wall St.: Dow off 37, Nasdaq up 22. Fed increase is less than some investors expected, but markets face uncertainty over potential for more hikes.
by Thomas S. Mulligan, *Los Angeles Times*, Home Edition, p. C1

The Fed's official statement Wednesday said the central bank believes the risks are "weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future." Many Wall Streeters believe that will mean at least two more quarter-point rate increases this year. Others, however, say rates may well be peaking now, assuming the economy slows and inflation remains subdued.

02/03/2000
WEDNESDAY'S MARKETS
Fed Decision Gets Mixed Reaction from Stocks—Treasury Move Stirs Bond Rally
by E.S. Browning, *Wall Street Journal*, p. C1

And while the Fed didn't explicitly say that it will raise rates again, that is what investors concluded from its comment that inflation remains the main threat to the economy.

02/03/2000
Fed Hikes Rates 0.25% Amid Concerns About Surging U.S. Economy
by Gerard Baker in Washington and Ed Crooks in London, *Financial Times*, USA Edition, p. 1

Amid concerns the robust U.S. economy could ignite inflation, the Federal Reserve on Wednesday raised two short-term interest rates to their highest level in more than four years and indicated that further tightening may be needed in the near future.

03/22/2000
TUESDAY'S MARKETS
Stocks and Bonds Shoot Higher Despite Rate Increase by the Fed
by Gregory Zuckerman, *Wall Street Journal*, p. C1

And despite investors' satisfaction that interest-rate increases must surely be winding down, the Fed hinted that more rate increases are likely this year.

03/22/2000
Fed Makes Expected Increase on Rates—Main Target Rises to 6%, More Action Promised; Markets Seem Unfazed
by Jacob M. Schlesinger, *Wall Street Journal*, p. A3

The Federal Reserve continued its slow but steady campaign to damp the economy with another small increase in interest rates and declared more action was likely this spring.

03/22/2000
Fed Raises Rates as Inflation Hedge; Markets Anticipated Quarter-Point Rise
by John M. Berry, *Washington Post*, Final Edition, p. E2

Federal Reserve officials, concerned that the nation's extraordinarily strong economic growth will eventually lead to higher inflation, raised short-term interest rates by a quarter-percentage point yesterday and indicated that more such moves are probable if growth doesn't slow to a more sustainable pace.

05/17/2000
Fed Raises Key Interest Rates; Policymakers Hint More Increases Will Follow Half-Point Boost
by John M. Berry, *Washington Post* , p. A1

The Federal Reserve raised its target for overnight interest rates by half a percentage point yesterday to slow headlong U.S. economic growth and keep inflation from rising. A statement explaining the action indicated that additional rate increases are likely in coming months.

05/17/2000
Fed Targets Inflation, Hikes Rate Half-Point; ECONOMY: Central bank hints at still more tightening. Banks quickly raise their prime rates, meaning consumers will soon feel the pinch.
by Thomas S. Mulligan, *Los Angeles Times*, Home Edition, p. A1

Escalating its campaign to preempt inflation, the Federal Reserve on Tuesday raised a benchmark interest rate by one-half of a percentage point and hinted that there may be more credit tightening ahead.

05/17/2000
Investors Shrug Off Fed Rate Rise, Push Blue Chips Up 126.79 Points
by E.S. Browning, *Wall Street Journal*, p. C1

The Fed hint of further rate increases, in the form of a warning that inflation remains a serious risk, came on top of its widely expected decision to raise its guideline short-term interest rates by half a percentage point.

05/17/2000
Fed Boosts Rates by One-Half Point, Warns That the Economy Isn't Slowing—Central Bank's Statement Indicates More Increases as Elections Approach
by Jacob M. Schlesinger, *Wall Street Journal*, p. A2

Looking ahead to their next meeting scheduled for June 27 and 28, the Fed's monetary policy committee said in a statement that it still believes the economy's "risks are weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future." That means more rate increases, possibly another half-point rise, are on the table, which could put Fed Chairman Alan Greenspan in the politically awkward position of continuing to raise borrowing costs as the November elections approach. Futures markets designed to predict upcoming Fed moves were betting late yesterday on a quarter-point move in June and then a half-point move by the November vote.

05/17/2000
Fed Tries To Rein in US Economy with Half Point Interest Rate Rise
by Gerard Baker, *Financial Times*, p. P1

In a statement, the FOMC attributed its decision to familiar concerns over growth in demand surpassing the growth in supply and implied further increases may be necessary in the near future.

06/29/2000
Market Savvy Fed Votes To Put Off 7th Straight Rate Hike; ECONOMY: The central bank implied it would raise interest rates in August unless it sees more signs of economic weakness.
by Thomas S. Mulligan, *Los Angeles Times*, Home Edition, p. C1

Thus the central bank implied it would resume raising borrowing costs at its Aug. 22 meeting unless it sees more definitive signs of economic weakness over the next two months.

06/29/2000
Fed Votes Not To Raise Interest Rates, for Now
by John M. Berry, *Washington Post*, Final Edition, p. E1

"Nonetheless, signs that growth in demand is moving to a sustainable pace are still tentative and preliminary" and the nation's labor markets remain very tight, the committee said, adding that "the risks continue to be weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future." That was the Fed's way of warning that inflation concerns could cause policymakers to move rates higher in coming months.

08/23/2000
Fed Holds Rates Steady, Issues Inflation Warning; ECONOMY: Many economists expect no further increases by the Federal Open Market Committee, which cited slowing growth in demand and improved productivity in its decision.
by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. C1

That murkiness was behind the Fed's warning Tuesday that the economic "risks continue to be weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future."

The warning signaled that the central bank is maintaining its "tightening bias," or tilting toward raising, rather than lowering, rates if it takes any further action at all.

08/23/2000
Market Skips Party on Fed News, Settles for Slim Gains
by E.S. Browning, *Wall Street Journal,* p. C1

Not everyone was so hopeful. The Fed indicated in a statement after its policy meeting that it might have to raise rates later to stave off inflation. "The risks," it said, "continue to be weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future."

11/10/2000
Analysts Predict Fed Will Leave Rates Alone
by John M. Berry, *Washington Post*, Final Edition, p. E1

Many analysts predict that the FOMC, the central bank's top policymaking group, will keep that "bias" in the statement it issues after its deliberations, and which many investors take as a hint at the Fed's possible future action. But some others believe economic growth has slowed enough that the FOMC may be ready to drop that bias in favor of a "neutral" statement saying the risks are now balanced. The issue is important because a shift to neutral probably would be taken by financial markets as a signal that rate cuts could be in offing, perhaps as soon as early next year.

11/16/2000
Fed Puts Rates On Hold Again; But Inflation Is Still Viewed as a Risk
by John M. Berry, *Washington Post*, Final Edition, p. E1

Some investors had been hoping the Fed officials would decide that those risks are balanced, which could be a first step toward reducing rates in coming months.

11/16/2000
Bond Prices Rally as Investors, Who See Chance of Future Cut, React to Fed's Decision on Rates
by Gregory Zuckerman and Steven Vames, *Wall Street Journal*, p. C20

The Fed left the fed-funds rate, or its target Fed funds overnight interbank rate, at 6.5%, and also kept in place its bias toward a "risk of heightened inflation pressures" and the higher interest rates that would be needed to fight such pressures. The Fed's decision was widely anticipated, but while some investors were disappointed the Fed kept its so-called bias tilted in favor of further rate increases others took heart by the wording in the Fed's statement accompanying its announcements, figuring the chances for a rate cut down the line had been raised.

12/20/2000
Fed Shifts to Worry Over Risk of Slump but Keeps Short-Term Rates Unchanged—Markets Now Expect Cuts at the End of January as Inflation Fear Fades
by David Wessel and Gregory Zuckerman, *Wall Street Journal*, p. A2

The Federal Reserve promised to throw a life preserver to the U.S., declaring that the risks of "economic weakness in the foreseeable future" exceed the risks of inflation. But it left short-term interest rates unchanged. Financial markets now expect the Fed to begin cutting rates at the end of January, and to reduce them at least one-half percentage point by spring.

12/20/2000
Fed Leans Toward a Future Rate Cut; Agency Says Economic Slowdown Poses Greater Threat Than Inflation
by John M. Berry, *Washington Post*, Final Edition, p. E1

The officials left interest rates unchanged for now but said the risk that economic growth will slow sharply is now greater than the risk that inflation will get worse. At their previous meeting, last month, they noted that growth had slowed but said inflation still posed the greater risk. That 180-degree swing in concern underscored the rapidity with which economists and Fed officials alike have been marking down their expectations about the immediate course of the economy. And to some analysts it suggested that the Fed officials could begin to cut rates as early as their next meeting, on Jan. 30-31.

12/20/2000

Treasury Prices Drop as Investors Absorb News of the Federal Reserve's New Stance on Rates

by Michael S. Derby, *Wall Street Journal*, p. C21

The Fed left its target for the federal-funds, or overnight bank, lending rate, unchanged at 6.50%. But its announcement of the decision suggested that the Fed is leaning toward lowering interest rates in the near future. It said that although there remains some potential for inflation, the risks to the economy now "are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

01/04/2001

Fed Unexpectedly Cuts Key Rate by Half-Point; ECONOMY: The central bank's aggressive action underscores concerns over a slowdown. Stock markets soar, with Nasdaq index posting a record gain.

by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. A1

As it customarily does, the Fed accompanied its rate reductions with an explanation. It said in a statement that it had acted "in light of further weakening of sales and production, and in the context of lower consumer confidence, tight conditions in some segments of financial markets and high energy prices sapping household and business purchasing power." And it left the door open for further cuts, saying that the risks "are weighted mainly toward conditions that may generate economic weakness."

01/04/2001

Fed Acts To Bolster Economy with an Unexpected Rate Cut; Wall Street Cheers Half-Point Move; Nasdaq Index Gains a Record 14.2%

by Steven Pearlstein, *Washington Post*, Final Edition, p. A1

The Federal Reserve yesterday cut short-term interest rates by half a percentage point—a dramatic move designed to bolster investor and consumer confidence and prevent the economy from slipping into recession. The Fed also hinted that further cuts may be in the offing.

01/04/2001

Two-Edged Sword: Fed's Surprise Move Sparks Market Rally, Sets Off New Jitters—Action Reflects Rising Fear of Recession, Pressure On Central-Bank Policy—Boost for Bush's Tax Cut?

by Jacob M. Schlesinger, Greg Ip, and Nicholas Kulish, *Wall Street Journal*, p. A1

By declaring that the "risks" in the economy remain "weighted mainly toward...economic weakness," the Fed also made clear it's ready to do still more. Financial markets that bet on future Fed moves were trading yesterday afternoon on the assumption of a quarter-point rate reduction at the central bank's two-day meeting on Jan. 30 and 31, and one more of the same size at the Fed's March meeting.

03/21/2001

Fed Delivers Rate Cut with a Hint of More to Come; ECONOMY: The half-point trim is intended to help revive growth, which has slowed to nearly zero. Analysts see the move as a refusal to accommodate Wall Street.

by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. A1

Among other things, officials promised to "monitor developments closely"—a code phrase that analysts said meant the central bank could cut rates again before the next meeting of its policymaking Federal Open Market Committee in mid-May.

03/21/2001

Economic Fix: As Fed Trims Rates, Other Forces Work To Dilute the Benefits—Consumer Debt, Slow Exports and Corporate Jitters Damp Jump-Start Bid—Markets Lose More Ground

by Greg Ip and Jacob M. Schlesinger, *Wall Street Journal*, p. A1

The Federal Reserve's move to lower short-term interest rates by half a percentage point brings rates down a total of 1.5 percentage points this year. And officials made clear yesterday they are prepared to do much more—possibly even before their next official rate-setting session May 15.

03/21/2001
Fed Rate Cut Leaves Wall St. Unsatisfied
by John M. Berry, *Washington Post*, Final Edition, p. A1

The FOMC cautioned in a statement that weak conditions in the manufacturing sector, where production has fallen and thousands of workers have been laid off, "could continue for some time." Increasing economic problems abroad, particularly in Japan, also pose "substantial risks" that could keep the U.S. economy soft for some time to come, the committee said. Because of these developments, the committee said, the risks the economy faces continue to be "weighted mainly toward conditions that may generate economic weakness in the foreseeable future." In other words, a solid recovery is not yet assured and more rate cuts may be needed.

04/19/2001
Bonds Rise on Fed's Surprise Interest-Rate Cut; Belief in More Trims Aids Short-Term Securities
by Gregory Zuckerman, *Wall Street Journal*, p. C16

Some bond traders seized on the wording of the Fed's rate announcement, which focused on the weakness of the U.S. economy and made it plain that the Fed is ready to keep cutting rates.

04/19/2001
Behind the Surprise: Half-Point Rate Cut Shows Balancing Act by Federal Reserve—Greenspan Pegs His Moves to the Economy but Keeps Close Eye on the Markets—Nasdaq Bounds 8.1% Higher
by Greg Ip and Jacob M. Schlesinger, *Wall Street Journal*, p. A1

Many economists think the Fed still has more work to do, which the central bank doesn't seem inclined to dispute. Its statement expressed concerns of continued "risks" of "economic weakness." The futures market that bets on Fed action is pricing in a Fed funds rate as low as 4% by July.

05/16/2001
THE NATION: Fed Cuts Key Rate Half a Point to 4%
by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. A1

The Federal Reserve cut its key interest rate another half percentage point to 4% on Tuesday and, contrary to what had been expected, left the door open for still more cuts aimed at getting the stumbling U.S. economy moving again.

05/16/2001
Fed Delivers Expected Rate Cut, but Investors' Reaction Is Muted
by E.S. Browning, *Wall Street Journal*, p. C1

The Fed did give investors just about all they could have hoped for: another half-percentage-point cut in its target for short-term interest rates and a hint that it will continue to reduce rates if the economy remains weak.

05/16/2001
Fed Makes 5th Cut in Rates This Year; Action to Date Is Most Aggressive Since '82
by John M. Berry, *Washington Post*, Final Edition, p. A1

The wording of the Fed's announcement gave no hint that the officials believe that economic growth is picking up. To the contrary, it signaled that the Fed is likely to cut rates again, though probably not before the next policymaking session in late June…The FOMC signaled in its statement that it will consider additional rate cuts, concluding that the risks facing the economy "are weighted mainly toward conditions that may generate economic weakness in the foreseeable future." But the statement omitted other language that has been used in recent months to prepare financial markets for a rate cut during the period between policymaking sessions.

06/28/2001
THE NATION: Fed Trims Key Rate a Quarter Point
by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. A1

The central bank's Federal Open Market Committee coupled the cut with a statement signaling it is ready to reduce rates further if economic troubles worsen. But both the action and the words seemed

considerably less emphatic than in past months, suggesting uncertainty about how much more is needed to spark a recovery.

06/28/2001

Fed Trims Interest Rates Again; Quarter-Point Reduction Disappoints Some Analysts

by John M. Berry, *Washington Post*, Final Edition, p. A1

Federal Reserve officials reduced their target for short-term interest rates by a quarter of a percentage point yesterday, the sixth rate cut of the year, as part of the central bank's effort to boost the country's anemic economic growth. They also left the door open to additional rate cuts by indicating in a statement that they still believe the risks to the economy "are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

06/28/2001

Financial Times

After five half-point interest rate cuts in five months, Wednesday's quarter-point move might disappoint some investors. But shifting to a quarter does not mean that the Fed's work is done: the statement maintains the bias towards cutting rates further.

06/29/2001

Fed Minutes Hint Cuts Are Nearly Over

by John M. Berry, *Washington Post*, Final Edition, p. E1

Wednesday's statement said the FOMC still found that the risk of further economic weakness outweighed the risk of inflation. In other words, as of this week the committee believes further rate cuts may be needed to stimulate the sluggish U.S. economy. But that does not necessarily mean there will be additional cuts. That will depend on the policymakers' assessment of the course of the economy when they meet next, on Aug. 21.

08/22/2001

Financial Times

The latest quarter point cut was justified by the continued weakness of the economy. Recent glimmers notwithstanding, the risks remain on the downside and the Fed maintains its policy bias to ease further.

08/22/2001

Recession Fears Prompt Fed To Cut Rates Again; ECONOMY: Panel trims benchmark a quarter point and leaves door open for another reduction to counter continuing weakness.

by Warren Vieth, *Los Angeles Times*, Home Edition, p. A1

The Federal Reserve cut interest rates Tuesday for the seventh time this year, warning that the economy may continue to weaken and signaling its willingness to ease rates even more to ward off a recession.

08/22/2001

Fed Again Reduces Key Rate; Quarter-Point Cut May Not Be Last

by John M. Berry, *Washington Post*, Final Edition, p. A1

Federal Reserve officials, concerned about the uncertain outlook for the U.S. economy amid a global slowdown in growth, lowered their target for short-term interest rates yesterday for a seventh time this year and left the door open for more cuts if needed...The Fed policymakers made it clear they are open to more cuts by saying that while "long-term prospects for productivity growth and the economy remain favorable," the committee believes "the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

08/22/2001

Fed Cuts Rates to Lowest Level Since '94—Quarter-Point Reduction Tied to Slump Overseas, Weak Business Climate

by Greg Ip, *Wall Street Journal*, p. A2

The Fed also indicated it was more likely to lower than raise rates in the future, saying risks were "weighted mainly toward...economic weakness" rather than inflation.

09/18/2001
ASSAULT ON AMERICA; ECONOMY: Fed's pre-emptive strike to shore up confidence
by Gerard Baker, *Financial Times*

> But the central bank also went out of its way to ensure there was no doubt that it was prepared to do much more to help out. Not only did it retain its policy "tilt"—the pro-forma statement that said the Fed sees the risks weighted more towards economic weakness than towards inflationary pressures...The maintenance of the "tilt" towards further easing was also strongly suggestive that interest rates are set to go lower still at the next scheduled meeting of the open market committee on October 2.

10/03/2001
Fed Trims Rates Again, Hints at Further Cuts
by John M. Berry, *Washington Post*, Final Edition, p. A1

> Federal Reserve policymakers, citing the damage caused by the recent terrorist attacks to the stalled U.S. economy, yesterday cut short-term interest rates for the ninth time this year and signaled that they may well reduce them again to help ease the coming financial pain...Fed officials also indicated yesterday that they are likely to further trim the federal funds rate, the interest rates financial institutions charge one another on overnight loans, perhaps as soon as their next meeting on Nov. 6. Even though that rate is already a full 4 percentage points lower than it was at the beginning of the year, the committee said "the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

11/07/2001
Key Fed Rate Cut to 40-Year Low; ECONOMY: In real terms, the half-point trim to 2% pushes the benchmark into negative territory.
by Peter G. Gosselin, *Los Angeles Times*, Home Edition, p. A1

> In what has become a familiar refrain, the central bank signaled it was ready to cut rates still further. "The risks are [still] weighted mainly toward conditions that may generate economic weakness," they said.

11/07/2001
Fed Lowers Rates for 10th Time This Year
by John M. Berry, *Washington Post*, Final Edition, p. A1

> Federal Reserve officials, clearly worried that the U.S. economy may be spiraling downward into recession in the wake of the Sept. 11 terrorist attacks, yesterday cut short-term interest rates for the 10th time this year and indicated they may trim them again if necessary...The FOMC also said that the risks facing the economy "are weighted mainly toward conditions that may generate economic weakness," the members' signal that they may cut rates again unless they see signs that the economy's downward momentum is slowing. Some analysts said such action is likely at the next FOMC meeting Dec. 11.

12/12/2001
With the Economy Still Fragile, the Fed Again Cuts Rates
by Richard W. Stevenson, *New York Times*, p. 1, col. 2

> The statement went on to use the Fed's code for a willingness to cut rates again, saying "the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

12/12/2001
Fed Slices Interest Rates to 1.75%, Leaves Door Open for More Cuts
by Greg Ip, *Wall Street Journal*, p. A2

> In a brief statement accompanying the move, policy makers said that "weakness in demand shows signs of abating, but those signs are preliminary and tentative." They said risks were still skewed to more economic weakness, suggesting an inclination to lower rates rather than raise them or leave them alone.

How Well Do Monetary Fundamentals Forecast Exchange Rates?

Christopher J. Neely and Lucio Sarno

In the last decade or so, exchange rate economics has seen a number of important developments, with substantial contributions to both the theory and the empirical understanding of exchange rate determination. Important developments in econometrics and the increasing availability of high-quality data have also stimulated a large amount of empirical work on exchange rates. While this research has improved our understanding of exchange rates, a number of challenges and questions remain. One of the most widely studied and still unanswered questions in this literature involves why monetary models of exchange rate determination cannot forecast much of the variation in exchange rates.

The monetary approach to exchange rate determination emerged as the dominant exchange rate model at the outset of the recent float in the early 1970s and remains an important exchange rate paradigm (Frenkel, 1976; Mussa, 1976, 1979; Bilson, 1978). However, Meese and Rogoff's (1983a) finding that monetary models' forecasts could not outperform a simple no-change forecast was a devastating critique of standard models and marked a watershed in exchange rate economics. Moreover, even with the benefit of 20 years of hindsight, evidence that monetary models can consistently and significantly outperform a naïve random walk is still elusive (e.g., see Mark and Sul, 2001; Rapach and Wohar, 2001a, 2001b; Faust, Rogers, and Wright, 2001).

This article reviews this puzzle and discusses several potential explanations for the consistent failure of monetary models to forecast much variation in nominal exchange rates. We present the essential elements of the monetary model in the next section and then discuss, in the third section, the key empirical studies examining the out-of-sample forecasting performance of the monetary model. The fourth section outlines possible explanations of the apparent failure of monetary model predictions and a final section briefly concludes.

Christopher J. Neely is a research officer at the Federal Reserve Bank of St. Louis. Lucio Sarno is a professor of finance in the Finance Group, Warwick Business School, University of Warwick, and a research affiliate of the Centre for Economic Policy Research (CEPR), London. The authors thank Nelson Mark and Jeremy Berkowitz for making programs and data available and Menzie Chinn, Nelson Mark, David Rapach, and Mark Wohar for helpful comments. Charles Hokayem provided research assistance.

THE MONETARY APPROACH TO EXCHANGE RATE DETERMINATION

In this section we describe the main features of the monetary approach to exchange rate determination in its flexible-price formulation (Frenkel, 1976; Mussa, 1976, 1979).[1]

The monetary approach starts from the definition of the exchange rate as the relative price of two monies and attempts to model that relative price in terms of the relative supply of and demand for those monies. In discrete time, monetary equilibria in the domestic and foreign country, respectively, are given by

$$m_t = p_t + \kappa y_t - \lambda i_t \quad (1)$$

$$m_t^* = p_t^* + \kappa^* y_t^* - \lambda^* i_t^*, \quad (2)$$

where m_t, p_t, y_t, and i_t denote the log-levels of the money supply, the price level, income, and the level of the interest rate, respectively, at time t; κ and λ are positive constants; asterisks denote foreign variables and parameters. In the monetary model, the real interest rate is exogenous in the long run and determined in world markets, because of the implicit assumption of perfect capital mobility.

Another building block of the monetary model is absolute purchasing power parity (PPP), which holds that goods-market arbitrage will tend to move the exchange rate to equalize prices in two countries. For example, if U.S. goods are more expensive than Mexican goods, U.S. and Mexican consumers will tend to purchase more goods in Mexico and fewer in the United States. The increased relative demand for Mexican goods will tend to make the peso appreciate with respect to the dollar and equalize the dollar-denominated prices of U.S. and Mexican goods. The monetary model assumes that PPP holds continuously, so that

$$s_t = p_t - p_t^*, \quad (3)$$

[1] For a more comprehensive discussion of the monetary model and other models of exchange rate determination, see Sarno and Taylor (2002, Chap. 4 and 5).

where s_t is the log-level of the nominal bilateral exchange rate (the domestic price of the foreign currency).

The domestic money supply determines the domestic price level and hence the exchange rate is determined by relative money supplies. Subtracting equation (2) from equation (1), solving for $(p_t - p_t^*)$, and inserting the result into equation (3) provides the solution for the nominal exchange rate:

$$s_t = (m_t - m_t^*) - (\kappa y_t - \kappa^* y_t^*) + (\lambda i_t - \lambda^* i_t^*), \quad (4)$$

which is the fundamental equation of the monetary model. The model is often simplified by assuming that the income elasticities and interest rate semi-elasticities of money demand are the same for the domestic and foreign country ($\lambda = \lambda^*$ and $\kappa = \kappa^*$) so that equation (4) reduces to

$$s_t = (m_t - m_t^*) - \kappa(y_t - y_t^*) + \lambda(i_t - i_t^*). \quad (5)$$

According to equation (5), an increase in the domestic money supply relative to the foreign money stock, for example, induces a depreciation of the domestic currency in terms of the foreign currency. In other words, the nominal exchange rate, s_t, increases. Conversely, a boost in domestic real income, *ceteris paribus*, creates an excess demand for the domestic money stock. In an attempt to increase their real money balances, domestic residents reduce expenditure and prices fall until money market equilibrium is achieved. Via PPP, the fall in domestic prices (with foreign prices constant) implies an appreciation of the domestic currency in terms of the foreign currency (a rise in the value of domestic currency in terms of foreign currency).

The model further assumes that the uncovered interest parity (UIP) condition holds

$$E_t\left(\Delta s_{t+1}\right) = (i_t - i_t^*), \quad (6)$$

where Δ is the first-difference operator, so that $\Delta x_t = x_t - x_{t-1}$ for any x, and $E_t(\Delta s_{t+1})$ denotes the market expectation of the change in the exchange rate.[2] The expected rate of depreciation of the domestic currency, Δs^e_{t+1}, can then be substituted for the nominal interest rate differential, $(i_t - i_t^*)$ in equation (5) to yield

$$s_t = (m_t - m_t^*) - \kappa(y_t - y_t^*) + \lambda E_t\left(\Delta s_{t+1}\right). \quad (7)$$

Using the identity $E_t(\Delta s_{t+1}) = E_t(s_{t+1}) - s_t$, equation (7) may in turn be rewritten as

$$s_t = (1+\lambda)^{-1}(m_t - m_t^*) - \kappa(1+\lambda)^{-1}(y_t - y_t^*) + \lambda(1+\lambda)^{-1} E_t\left(s_{t+1}\right). \quad (8)$$

By iterating forward in equation (8), the rational expectations solution to (7) may be written as

$$s_t = (1+\lambda)^{-1} \sum_{i=0}^{\infty} \left(\frac{\lambda}{1+\lambda}\right)^i E_t\left[(m_{t+i} - m_{t+i}^*) - \kappa(y_{t+i} - y_{t+i}^*)\right], \quad (9)$$

where $E_t[\cdot]$ denotes the mathematical expectation conditional on the information set available at time t.[3] It is well known from the rational expectations literature, however, that equation (9) represents only one solution to (7) from a potentially infinite set. In general, given the exchange rate determined according to equation (9), say s_t^E, (7) has multiple rational expectations solutions conforming to

$$s_t = s_t^E + B_t, \quad (10)$$

where the rational bubble term B_t satisfies

$$E_t\left[B_{t+i}\right] = \lambda^{-1}(1+\lambda)B_t. \quad (11)$$

Therefore, s_t^E simply represents the rational expectations solution to the monetary model in the absence of rational bubbles. Rational bubbles represent significant departures from the fundamentals of the model that would not be detected in a specification such as (5). Thus, testing for the presence of bubbles can be interpreted as an important specification test of the model (Meese, 1986).

Assumptions of the Monetary Model

Although the simplicity of the flexible-price monetary model is very attractive, this simplicity requires many assumptions. Open economy macroeconomics is essentially about six aggregate markets: goods, labor, money, foreign exchange, domestic bonds (i.e., non-money assets), and foreign bonds. The monetary model concentrates, however, directly on equilibrium conditions in only one of these markets, the money market. This is implicitly achieved in the following fashion. By assuming perfect substi-

[2] UIP states that risk-neutral arbitrage will equalize the expected return on a foreign investment—approximately $(E_t(\Delta s_{t+1}) + i_t^*)$—and the return on a domestic investment (i_t).

[3] Moving from equation (8) to equation (9) requires writing the expression for s_{t+1} in terms of $E_{t+1}(s_{t+2})$ implied by (8), taking expectations, substituting the result for $E_t(s_{t+1})$ in (8), and then repeating the process for $E_{t+2}(s_{t+3})$, $E_{t+3}(s_{t+4})$, etc.

tutability of domestic and foreign assets, the domestic and foreign bond markets essentially become a single market, reducing the number of markets to five. The exchange rate adjusts freely to equilibrate supply and demand in the foreign exchange market. Perfectly flexible prices and wages likewise equilibrate supply and demand in the goods and labor markets. Thus, three of the five remaining markets are cleared. Recalling Walras's law, according to which equilibrium in $n-1$ markets of an n-market system implies equilibrium in the nth market, equilibrium of the full system in the model is then determined by equilibrium conditions for the money market. The flexible-price monetary model is thus, implicitly, a market-clearing general equilibrium model in which continuous PPP among national price levels is assumed (Sarno and Taylor, 2002, Chap. 4).

Sticky-Price Monetary Models

The very high volatility of real exchange rates during the 1970s float cast serious doubts on the assumption of continuous PPP and inspired the development of further classes of models, including sticky-price monetary models and equilibrium models.[4]

The sticky-price monetary model, due originally to Dornbusch (1976), allows short-term overshooting of the nominal and real exchange rates above their long-run equilibrium levels. In this model, it is assumed that there are "jump variables" in the system (exchange rates and interest rates) compensating for stickiness in other variables, notably goods prices. Consider the effects of a cut in the nominal domestic money supply. Since goods prices are sticky in the short run, this implies an initial fall in the real money supply and a consequent rise in interest rates to clear the money market. The rise in domestic interest rates then leads to a capital inflow and an appreciation of the nominal exchange rate. Investors are aware that they are artificially forcing up the value of the domestic currency and that they may therefore suffer a foreign exchange loss when the proceeds of their investment are used to repay liabilities in foreign currency. Nevertheless, as long as the expected foreign exchange loss (the expected rate of depreciation of the domestic currency) is less than the known capital market gain (the interest rate differential), risk-neutral investors will continue to borrow abroad to buy domestic assets. A short-run equilibrium is achieved when the expected rate of depreciation is just equal to the interest rate differential, i.e., when UIP holds. Since the domestic currency must be expected to depreciate because of the interest rate differential, the domestic currency must have appreciated beyond its long-run, PPP equilibrium. In the medium run, however, domestic prices begin to fall in response to the fall in the money supply. This alleviates pressure in the money market (the real money supply rises), and domestic interest rates start to decline. The exchange rate then depreciates slowly toward long-run PPP. Thus, this model can explain the apparent paradox that the currencies of countries with relatively higher interest rates tend to depreciate: the initial rise in the interest rate induces a sharp exchange rate appreciation, followed by a slow depreciation as prices adjust, which continues until long-run PPP is satisfied.

Nevertheless, it should be clear that, regardless of whether one assumes that prices are flexible or sticky, the traditional flexible-price monetary model and its sticky-price formulation imply exactly the same fundamental equation for the exchange rate, which is of the form (5). We now turn to the empirical evidence on the performance of the monetary model in forecasting exchange rates.

FORECASTING EXCHANGE RATES WITH MONETARY MODELS

The move to floating exchange rates in the 1970s spawned a wealth of theoretical efforts to explain their observed high volatility. The monetary models discussed in the previous section were among the most popular and intuitively appealing. It was natural to examine the empirical fit and forecasting ability of these models. This section selectively reviews the long literature attempting to use monetary models to forecast exchange rates.[5]

---

[4] Equilibrium exchange rate models, due originally to Stockman (1980) and Lucas (1982), analyze the general equilibrium of a two-country model in a representative agent, utility maximizing framework with sound microfoundations. Equilibrium models may be viewed as an extension or generalization of the flexible-price monetary model that allows for multiple traded goods and real shocks across countries. These models are not amenable to direct econometric testing or to the formulation of models designed to forecast exchange rates because they are based on utility functions that cannot be directly estimated. (Rather, researchers have sought to test the broad rather than specific implications of this class of models for exchange rate behavior.) Similar reasoning applies, at least at the present stage, to the literature on "new open economy macroeconomics" (see Lane, 2001, Sarno, 2001, and the references therein).

[5] This paper focuses on forecasting exchange rates with monetary models. There are many nonmonetary models available, however. Fair (1999) uses a nonmonetary macro model; Clarida and Taylor (1997) and Clarida, Sarno, Taylor, and Valente (2003) use models based on the term structure; and Evans and Lyons (1999) use order flow models to explain exchange rate changes. Cheung, Chinn, and Pascual (2002) examine the performance of the most recent set of linear models.

Meese and Rogoff (1983a and 1983b)

Meese and Rogoff (1983a, 1983b)—hereafter MR—conducted the seminal work in the use of monetary models to forecast the exchange rate. Their procedure was straightforward: They regressed the log of exchange rates on various combinations of relative macroeconomic variables typically included in the exchange rate models of the 1970s.[6] The basic prediction equation was as follows:

(12)
$$s_{t+k} = a_0 + a_1(m_t - m_t^*) + a_2(y_t - y_t^*) + a_3(i_t - i_t^*) + a_4(\pi_t^e - \pi_t^{e*}) + a_5 tb_t + a_6 tb_t^* + u_t,$$

where s_t, m_t, y_t, i_t, π_t^e, and tb_t are the logs at time t of the exchange rate, domestic (U.S.) money supply, output, interest rates, expected inflation, and the trade balance. Asterisks denote foreign variables. MR interpreted exchange rate models, such as the Frenkel-Bilson, Dornbusch-Frankel, and Hooper-Morton models, as implying different sets of restrictions on the coefficients in the regression (Hooper and Morton, 1982). As is the case with most estimation of macroeconomic models, little effort was made to explicitly map the model to the functional form and estimation procedure.

The data were monthly observations from March 1973 through June 1981. MR estimated the models on in-sample periods by several techniques, including ordinary least squares (OLS), generalized least squares (GLS) (to correct for serial correlation in the errors), and Fair's (1970) instrumental variables (IV) (to correct for simultaneous equations bias).[7] To allow the out-of-sample forecast coefficients to change, rolling regressions with fixed sample sizes were used. That is, coefficients were initially estimated using data until November 1976, then 1-, 3-, and 12-month forecasts were constructed. To construct the next set of forecasts, the next month of data (December 1976) was added, the first month of data was dropped, and the coefficients were re-estimated. For the exercises in which future values of the independent variables were needed to construct forecasts, MR provided the models with actual future values of the independent variables—instead of forecasting them—to give the monetary model the best possible chance of forecasting well.[8]

MR used both in-sample model evaluation criteria, such as the R^2, and out-of-sample criteria, such as the comparison of the root-mean-squared error (RMSE) of the model's forecast with that of a benchmark forecast, the driftless random walk. Many of the estimated models fit the in-sample data well. In-sample evaluation techniques, which permit the use of all the data available to the researcher, provide more precise estimates of statistics of interest and therefore have greater *power* to reject the null hypothesis of no predictability of the exchange rate.[9] The advantage of out-of-sample evaluation procedures is that they implicitly test the stability of the estimated coefficients and therefore provide a more stringent and realistic hurdle for models to overcome.

The main conclusion of the MR paper was that none of the structural exchange rate models were able to forecast out-of-sample better than a naïve no-change forecast by mean-squared error (MSE) and mean absolute error (MAE) criteria. There was some evidence of predictability at longer horizons, but—given the massive failure at short horizons—this did not receive much attention.

Econometric Problems

The MR exercise had a number of econometric problems, many of which they recognized and attempted to mitigate with variations on their procedures. First, because the explanatory variables were all endogenous—determined within the economic system—the estimated coefficients in the equations surely suffered from simultaneous equations bias. That is, even with an arbitrarily large amount of data, the coefficient estimates would not converge to any structural parameters. MR (1983b) attempted to correct for this problem with IV estimation and an in-sample grid search over possible parameter values. The IV estimation did not help and an in-sample grid search constituted unconvincing evidence. Because the benchmark

[6] MR also estimated univariate models of exchange rate changes and vector autoregressions, employing all the variables in equation (12). These models were also unsuccessful, however, and this paper focuses on efforts with monetary models.

[7] MR were very thorough in checking the robustness of their results to changes in procedures; because of space constraints in this paper, it is difficult to list all their permutations of models, estimation methods, and data.

[8] Faust, Rogers, and Wright (2001) have recently shown that real-time, Federal Reserve forecasts of future independent money and output variables actually generate better forecasts of the future exchange rate than do actual future values of the independent variables.

[9] The *power function* of a statistical test is the probability of rejecting the null hypothesis, conditional on the true data-generating process. The *size* of a test is the power when the null hypothesis is true.

no-change prediction is nested within the model, some combination of parameters must perform at least as well as the no-change model within the sample. A model with all zeros for coefficients, for example, will perform exactly as well as the no-change forecast. And, in practice, some combination of coefficients will almost certainly outperform the no-change forecast, in-sample.

It is unclear, however, why biased coefficients would be a problem for a forecasting exercise. If the covariance matrix of the structural errors is homoskedastic and stable over time, forecasts from biased coefficients would be superior to those from structural parameters.

Second, problems with the persistence of the variables in the regression were not dealt with in a convincing way. (See the boxed insert on persistence and cointegration.) In particular, because the dependent variable, the log exchange rate, in the predictive regression was probably I(1) but not cointegrated with any combination of the independent variables, the error term was probably I(1) and the coefficient estimates were inconsistent and thus meaningless for forecasting.

More Negative Results

The very strong negative results of the MR study spawned an enormous amount of subsequent research that varied econometric techniques or the information set to try to rescue the ability of fundamental models—or any models—to forecast exchange rates. For example, Wolff (1987, 1988), Schinasi and Swamy (1989), and Canova (1993) used time-varying coefficients in an effort to compensate for instability in the model. Other authors used expanded information sets or different functional forms for the empirical work (Meese and Rose, 1990). Such efforts, however, proved immediately unsuccessful or were subsequently shown to be fragile to minor changes in technique or the data. The MR result that monetary fundamentals do not help predict exchange rates remained conventional wisdom.

Resurrecting the Monetary Approach: Mark (1995)

Some progress was made, however, in the 1990s. Mark (1995) and Chinn and Meese (1995) focused on neglected aspects of the problem, primarily the possibility of better long-run predictability, but also the finite sample properties of the test statistics and the estimation procedure implied by the theoretical model.

Mark (1995) considers an expression relating the change in the exchange rate to the deviation of the exchange rate from a linear combination of relative money and relative output, which is called the *fundamental* value of the exchange rate. Essentially Mark exploits the monetary model equation (5), assuming that $\kappa = 1$ and the interest differential is equal to zero, so that the fundamentals term is

$$f_t = \left[\left(m_t - m_t^*\right) - \left(y_t - y_t^*\right)\right]. \tag{13}$$

In this model, the difference between the current fundamentals and the current exchange rate—called the *error correction term* $(f_t - s_t)$—determines the k-period-ahead change in the exchange rate:

$$s_{t+k} - s_t = \alpha_k + \beta_k\left(f_t - s_t\right) + \upsilon_{t+k,t}. \tag{14}$$

Using quarterly data on U.S. dollar (USD) exchange rates with Canada (CAD), Germany (DEM), Japan (JPY), and Switzerland (CHF) from 1973:Q2 to 1991:Q4, Mark computed the forecast regression in (14) over horizons of 1, 4, 8, 12, and 16 quarters. Out-of-sample forecasts began in 1984:Q1.

As did MR, Mark (1995) relied on both in-sample and out-of-sample evidence to assess the degree of predictability in his model. In-sample evidence included both the statistical significance of the estimated coefficients, β_k, and the R^2s of the regressions at various horizons, R_k^2. Out-of-sample evidence focused on the RMSE provided by the forecasts from the estimated model versus those from a driftless random walk, the same out-of-sample benchmark as used by MR. Mark (1995) examined the ratio of these RMSEs, OUT/RW$_k$ (also called the Theil U statistic), as well as the statistical significance of their difference, the DM statistic (Diebold and Mariano, 1995).[10]

In evaluating the statistical significance of the results, Mark (1995) confronted some of the same econometric problems that beset MR. Two of the complications are well known: (i) Because the independent variable $(f_t - s_t)$ is highly autocorrelated, the coefficients α_k and β_k would be biased in finite samples (Mankiw and Shapiro, 1986; Stambaugh, 1986); and (ii) for forecast horizons (k) greater than one period, the overlapping forecast errors would have autocorrelation of at least degree $k - 1$, com-

[10] Mark (2001) provides an easily readable discussion of exchange rate forecasting issues.

plicating the construction of functions of the forecast errors, like coefficient standard errors.

To overcome these difficulties, Mark (1995) relied on *nonparametric bootstrapping* to determine the statistical significance of his in-sample and out-of-sample results. Nonparametric bootstrapping is a method of simulating the distribution of statistics with the distribution of actual errors estimated by the model—rather than pseudo-random errors from a normal (or other) distribution—under some assumption about how the data were generated (Berkowitz and Kilian, 2000). In this case, a bootstrapping exercise calculates how often an economy in which there was no predictability would produce as much support for predictability as found in actual data. In other words, the question is: Do the real data look like they might have been generated by a model in which there is no predictability?

Mark (1995) assumed a null data-generating process (DGP) where the exchange rate change is simply a constant plus an error term and the error correction term (ECT) $(f_t - s_t)$ follows an autoregressive process of order P,

$$\Delta s_t = a_0 + \varepsilon_{1,t}, \tag{15}$$

$$\left(f_t - s_t\right) = b_0 + \sum_{j=1}^{P} b_j \left(f_{t-j} - s_{t-j}\right) + \varepsilon_{2,t}. \tag{16}$$

Note that (15) implies that the change in the exchange rate is unforecastable. Mark chose the parameters of the DGP to match the actual data and then constructed bootstrapped distributions for the test statistics as follows:

1. Estimate the null DGP, which is described by equations (15) and (16).
2. Draw 2000 error samples of size 76 from the estimated null DGP covariance matrix.
3. Use the errors to compute 2000 series of s_t and $f_t - s_t$, from equations (15) and (16).
4. Run the predictive regression, equation (14), to obtain estimates of β_k, its t statistics, and the out-of-sample RMSEs from the estimated models and the benchmark no-change prediction.

Table 1 presents selected results from Mark's (1995) exercise with significance levels generated from a DGP described by (15) and (16). Out-of-sample forecasts were evaluated against the benchmark of a driftless random walk—no change in the exchange rate. Mark (1995) concluded that evidence of predictability, including β_ks, adjusted β_ks, t statistics, and R^2s, increases with the forecast horizon and that there is evidence of statistically significant forecastability at horizons of 12 and 16 quarters for the DEM and CHF. In the German case, for example, the β_1 is 0.035 and the β_{16} is 1.324. The t statistics (p values) likewise rise (fall) with k, except for Canada. The p values for the German β_1 and β_{16} t statistics are 0.291 and 0.038, respectively. Likewise, the strongest out-of-sample evidence for predictability is at the longest horizons. The OUT/RW statistics—which are less than 1 when the monetary forecasting regression provides lower RMSEs than the no-change forecast—show that the monetary model beats the benchmark at every horizon for the CHF and the JPY and at the 12- and 16-quarter horizons for the DEM. In the latter case, the regression's RMSE is about half that of the no-change forecast at the 16-quarter horizon.

Unpredictability Strikes Back

Mark's (1995) innovative use of the bootstrap solved a number of econometric problems and appeared to show that there was greater power to predict exchange rates at long horizons than at short horizons. And his conclusions were largely buttressed by those of Chinn and Meese (1995), who investigated many of the same issues and used a wider variety of explanatory variables, including trade balance, the relative price of tradeables/nontradeables, interest rates, and inflation, as well as nonparametric methods. Chinn and Meese (1995) found that their fundamental-based error-correction models outperformed the random walk model for long-term prediction horizons.

Soon, however, other researchers such as Berkowitz and Giorgianni (2001) and Kilian (1999) began to criticize Mark's (1995) methodology and the resultant conclusions. Berkowitz and Giorgianni (2001) focused on how Mark's (1995) assumptions about the long-run behavior of the data series influenced the evidence of predictability. Kilian (1999) looked more carefully at the form of the assumed DGP and the robustness of the results to changes in the sample. Both criticisms focused on a disadvantage of bootstrapping and other simulation procedures: The results can depend crucially on the assumed DGP.

Berkowitz and Giorgianni (2001)—hereafter BG—pointed out that Mark's (1995) DGP—equations (15) and (16) in this paper—implicitly assumed that the exchange rate and the macroeconomic fundamentals were cointegrated, meaning that while each of the series $\{s_t, f_t\}$ might be individually I(1), a linear combination of them is stationary, or I(0).

Table 1

Mark's (1995) Results Using the DGP in Equations (15) and (16) (1973:Q2–1991:Q4)

Country	Horizon	Beta	Adj-Beta	t(20)	p Value	R^2	p Value	OUT/RW	p Value	DM(20)	p Value
Canada	1	0.040	0.029	3.051	**0.070**	0.059	**0.058**	0.998	0.181	0.061	0.184
	4	0.155	0.109	2.389	0.183	0.179	**0.090**	1.119	0.537	–1.270	0.472
	8	0.349	0.258	2.539	0.215	0.351	**0.065**	1.145	0.388	–1.036	0.361
	12	0.438	0.317	1.961	0.340	0.336	0.150	1.436	0.550	–1.916	0.531
	16	0.450	0.286	1.542	0.443	0.254	0.305	1.699	0.615	–2.596	0.542
Germany	1	0.035	0.011	1.836	0.291	0.015	0.419	1.015	0.340	–0.932	0.403
	4	0.205	0.106	2.902	0.181	0.104	0.267	1.037	0.289	–1.345	0.506
	8	0.554	0.363	3.487	0.191	0.265	0.178	1.002	0.226	–0.027	0.225
	12	0.966	0.676	6.329	**0.069**	0.527	**0.060**	0.796	0.109	4.246	**0.058**
	16	1.324	0.955	9.256	**0.038**	0.762	**0.015**	0.524	**0.036**	8.719	**0.045**
Japan	1	0.047	0.012	1.396	0.398	0.020	0.332	0.988	0.248	1.571	0.137
	4	0.263	0.132	2.254	0.278	0.125	0.205	0.928	0.210	2.302	0.121
	8	0.575	0.315	3.516	0.209	0.301	0.126	0.819	0.170	3.096	0.109
	12	0.945	0.564	4.889	0.152	0.532	**0.036**	0.712	0.149	3.319	0.146
	16	1.273	0.790	4.919	0.169	0.694	**0.011**	0.574	0.121	5.126	0.157
Switzerland	1	0.074	0.044	2.681	0.125	0.051	**0.096**	0.997	0.266	0.066	0.282
	4	0.285	0.167	3.248	0.148	0.180	**0.091**	0.981	0.256	0.218	0.265
	8	0.568	0.336	4.770	**0.095**	0.336	**0.077**	0.917	0.219	0.703	0.240
	12	0.837	0.509	8.013	**0.024**	0.538	**0.026**	0.738	0.122	2.933	0.135
	16	1.086	0.672	17.406	**0.001**	0.771	**0.001**	0.411	**0.026**	9.650	**0.071**

NOTE: The table was constructed using programs and data supplied by Nelson Mark. The null DGP constructs the exchange rate as a random walk with drift, the error correction term, $(f_t - s_t)$, is constructed to follow an AR(p) process, and errors to equations (15) and (16) are drawn with nonparametric bootstrapping. The benchmark for out-of-sample forecast comparison is the driftless random walk, a no-change forecast. "Beta" denotes the estimate of β_k from equation (14); "Adj-Beta" denotes the estimate of β_k adjusted for endogenous regressor bias; "t(20)" is the t statistic computed using a 20-period window for Newey-West corrected standard errors; "R^2" is the R^2 of the forecast equation (14); "OUT/RW" is the ratio of the forecast RMSE to the RMSE of the no-change benchmark; "DM(20)" is the Diebold-Mariano statistic, computed using a 20-period window for construction of the covariance matrix for the test of equality between the RMSEs of the regression forecast and the no-change benchmark forecast. Boldface p values denote results significant at the 10 percent level.

In other words, even if the difference between f_t and s_t is nonstationary in the real data, estimation of equation (16) will tend to generate data in which the difference between f_t and s_t is stationary. The generated exchange rate, s_t, cannot diverge very far from the generated macroeconomic fundamental, f_t. Ex ante, it is not obvious that cointegration is an important issue, as cointegration is neither a necessary nor a sufficient condition for fundamentals to predict exchange rate changes. However, in this case, BG argue that the distribution of the test statistics depends on whether there is cointegration or not.[11] If f_t and s_t are not cointegrated in the real data, then the critical values generated under the assumption of cointegration will be incorrect. The critical values will be incorrect because the forecasting regression,

[11] Berben and van Dijk (1998) derive the asymptotic distributions of the estimator of the regression parameter and its t statistic, under the null hypothesis of no cointegration. They find that the distribution does not depend on the forecast horizon; long-horizon regressions have no power advantages in testing for cointegration. Their analysis shows that Mark's (1995) results can be at least partly explained by his assumption of cointegration.

A SHORT PRIMER ON PERSISTENCE AND COINTEGRATION

A data series is said to be *stationary* if neither its mean nor any of its autocovariances depend on the date t. An implication of this is that the series will tend to be mean-reverting, to tend to return to its expected value when it departs from it. The longer such a series takes to return to its mean, the more *persistent* it is said to be. A series that is infinitely persistent will never tend to return to any mean value—its expected value doesn't exist; such a series is *nonstationary*.[1] Note that if a variable X_t has zero mean and is uncorrelated at all leads and lags—i.e., it is *white noise*—then X_t is stationary. If we define Y_t as the cumulated sums of X_t,

$$\text{i.e., } Y_t = \sum_{i=0}^{t} X_i,$$

then, as $t \to \infty$, the variance of Y_t becomes infinite, Y_t does not tend to revert to any value, so it is nonstationary. Y_t is also said to be I(1), *integrated of order one*—because it is the partial sum (or integral) of the X_is—and X is referred to as I(0). Y_t is also sometimes called a *random walk* because changes to it are unpredictable (random) from other information, $Y_t = Y_{t-1} + X_t$, $E(\Delta Y_t) = 0$. Note that if X_t had a constant non-zero mean, that is $X_t = \delta + \varepsilon_t$, where ε_t is white noise, then Y_t would be a *random walk with drift*: $Y_t = Y_{t-1} + \delta + \varepsilon_t$, $E(\Delta Y_t) = \delta$.

If at least one linear combination of I(1) variables is I(0), the variables are said to be *cointegrated*. Economic theory often implies that a linear combination of variables will be stationary. For example, one version of purchasing power parity says that exchange rates should be cointegrated with relative price levels—assuming that both are I(1) variables.

The persistence and cointegration of variables can have important implications for specifying regressions. For example, a regression equation only makes sense if the dependent and independent variables can be written so that the error is I(0). This requires that either both sides are I(0) or some combination of them is I(0). If the error term is I(1), its mean does not exist, coefficient estimates will be inconsistent, and forecasts will be biased. Generally, regressions are valid only if coefficients exist that make the error term stationary (Phillips, 1986).

A bivariate vector autoregression with a cointegrating relation between I(1) variables can be written in vector-error correction (VECM) form as follows:

$$\Delta s_t = \lambda_1 (f_{t-1} - s_{t-1}) + \sum_{i=1}^{p} a_i \Delta s_{t-i} + \sum_{i=1}^{p} b_i \Delta f_{t-i} + \mu_1 + \varepsilon_{1,t}$$

$$\Delta f_t = \lambda_2 (f_{t-1} - s_{t-1}) + \sum_{i=1}^{p} c_i \Delta s_{t-i} + \sum_{i=1}^{p} d_i \Delta f_{t-i} + \mu_2 + \varepsilon_{2,t}.$$

Under the null hypothesis of no cointegration between the variables, the individual variables are I(1), and a VAR in differences is appropriate. Such a VAR implies that $\lambda_1 = 0$ and $\lambda_2 = 0$. To test this hypothesis, Horvath and Watson (1995) suggest a Wald test of the null that $\lambda_1 = 0$ and $\lambda_2 = 0$ in the above representation.

Unfortunately, it is difficult to tell the difference between variables that are I(1)—having no mean—and those that are I(0) but highly persistent, taking a long time to return to their mean. Therefore, resolving the question of whether fundamentals are cointegrated with exchange

[1] If the expected value of a series does not exist, it is nonstationary. But nonstationary series can have expected values.

(*Continued on p. 59*)

(14), is almost a *spurious regression* (see the boxed insert): because, as the forecast horizon, k, increases, the change in the exchange rate, Δs_{t+k} in (14), becomes more persistent and—if there is no cointegration between s_t and f_t—the independent variable $(f_t - s_t)$ is I(1). When both sides of the forecasting equation are highly persistent, it approaches a *spurious regression* in which estimated coefficients falsely appear to be statistically significant. More generally, the distribution of the estimated coefficient from equation (14) will depend on the degree of persistence in the regressor $(f_t - s_t)$. If the null DGP fails to model the persistence of the regressor $(f_t - s_t)$ correctly, then the critical values of the forecasting statistics will be wrong and the inference drawn from the test might be incorrect.

Table 2 presents the results of a Horvath-Watson test (see the boxed insert) conducted by BG for cointegration of the exchange rate and macro fundamentals. BG were able to reject the null of no

(*Continued from p. 58*)
rates is more difficult than one might think. In fact, it can be shown that stationary and nonstationary processes are *observationally equivalent* (Blough, 1992; Faust,1993). In other words, there is a nonstationary process that will behave arbitrarily close to any given stationary process, and vice versa. In practice, it is frequently the case that a series of interest could be consistent with either stationarity or nonstationarity.

The problem of *spurious regression* occurs when an I(1) variable is regressed on an unrelated I(1) variable. Spurious regressions, investigated by Granger and Newbold (1974) and analyzed further by Phillips (1986), are regressions that relate independent random walks and produce apparently statistically significant—but inconsistent and meaningless—coefficients. Intuitively, spurious regressions occur because the moment matrix of the regressors ($X'X$) doesn't converge to anything and so becomes arbitrarily large, making the reported standard error of the coefficient too small and permitting false rejections of the hypothesis that the coefficient is equal to zero.

Even a variable that is not I(1) can cause problems in a regression if it is *persistent*—highly autocorrelated. It has long been understood that if the independent variable in a regression is persistent, the coefficient estimates in that regression will be biased. Marriott and Pope (1954) and Kendall (1954) studied the phenomenon in the autoregressive case and Mankiw and Shapiro (1986) and Stambaugh (1986) extended the results to multiple-equation models. To understand why persistence creates finite sample bias, consider the case of an AR(1) process: $y_t = \beta y_{t-1} + \varepsilon_t$, where $y_0 = 0$ and $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t) = 0$. The OLS estimator is given by

$$\hat{\beta} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1} y_{t-1}} = \frac{\sum_{t=1}^{T} y_{t-1}\left(\beta y_{t-1} + \varepsilon_t\right)}{\sum_{t=1}^{T} y_{t-1} y_{t-1}},$$

and the difference between $\hat{\beta}$ and β is given by:

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

With persistent regressors, the expectation of $\hat{\beta} - \beta$ is not equal to zero because there is positive correlation between the product $y_{t-1}\varepsilon_t$ and y_{t+j}^2 for $j \geq 0$. In other words, if the product of the regressor (y_{t-1}) and the error term (ε_t) is large and positive, then they are probably both large and of the same sign. In this case, near-term future values of the series (y_t) will tend to be far from the unconditional mean and their squares will be large too. Conversely, if $y_{t-1}\varepsilon_t$ is negative, then y_t will tend to be closer to its mean (0) than y_{t-1}, and so y_t^2 will tend to be small. Thus an estimator in a regression with a persistent regressor will be biased in finite samples.

cointegration for only one rate, the CHF. Kilian's results were even more negative toward the cointegration hypothesis; his test failed to reject the null of no cointegration for any exchange rate.[12] Unfortunately, it is often impossible to reject the null of no cointegration, even if $(f_t - s_t)$ is stationary but persistent. To better evaluate the balance of the evidence, Kilian adapts an idea of Rudebusch (1993) to weigh the balance of evidence for and against cointegration.[13] Kilian finds that the evidence favors the cointegration hypothesis for the CHF, was indeterminate for CAD, and favored the null of no cointegration for the DEM and JPY. Nevertheless, Kilian (1999) concludes that the data are potentially consistent with either the assumption of cointegration between s_t and f_t or the contrary, no cointegration. He notes that even if the series are cointegrated, the ECT $(f_t - s_t)$ reverts to its mean very slowly.

Because the Horvath-Watson test results imply that Mark's (1995) assumption of cointegration might be incorrect and because this assumption might influence the distribution of test statistics, BG reexamined the forecastability question without the cointegration assumption. In particular, BG conducted two bootstrapping experiments to study the behavior of the system under alternative assumptions about the DGP. Their first model assumed that

[12] Kilian's (1999) Horvath-Watson results might have been different because his sample and estimation methods were different from BG's.

[13] Rudebusch (1993) examines whether one can differentiate the short-run persistence properties under the best stationary model and the best nonstationary model.

Table 2

Results From a Horvath-Watson Test for Cointegration (1973:Q2–1994:Q4)

Country	λ_1	t Statistic	λ_2	t Statistic	Wald statistic	Lag length
Canada	0.031	1.60	–0.016	–1.48	5.801	2
Germany	0.033	1.11	0.001	0.11	1.233	2
Japan	0.057	1.49	–0.002	–0.25	2.442	3
Switzerland	0.079	1.98	–0.014	–1.78	7.842	2

NOTE: This table is excerpted from Table 7 of BG. It shows the results of a Horvath-Watson test for cointegration. Wald statistics exceeding the 10 percent critical value of 6.63 or the 5 percent critical value of 8.47 reject the null of no cointegration.

the exchange rate is a random walk with drift—as did Mark (1995)—and that macro fundamentals (f_t) follow an AR(3) process:

$$\Delta s_t = a_0 + \varepsilon_{1,t}, \tag{17}$$

$$f_t = b_0 + \sum_{j=1}^{3} b_j f_{t-j} + \varepsilon_{2,t}. \tag{18}$$

This first model did not assume cointegration and, in generating data, the covariance between the error terms $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ was set equal to zero. The exchange rate and macro fundamentals were independent by construction.

Table 3, which is excerpted from Table 4 in BG, shows the results of Mark's forecasting exercise with three changes: (i) p values were calculated with the DGP described in (17) and (18); (ii) parametric bootstrapping was used in place of nonparametric bootstrapping to generate data; and (iii) the out-of-sample benchmark included a drift term.[14] BG find that many of the DEM statistics—denoted by shaded boxes—are no longer significant.[15] Only the CHF shows much evidence of predictability.

Table 4—excerpted from Table 5 in BG—is constructed in exactly the same way as Table 3, except that it extends the sample by three years to the end of 1994. With this change, there is now no evidence of predictability at the 5 percent level, even at long horizons for any exchange rate except the CHF.[16] However, there is now evidence of predictability in the OUT/RW statistics at shorter horizons for the CHF.

The second BG model was an unrestricted vector autoregression (VAR) for the pair $\{s_t, f_t\}$. BG considered this model, which permitted but did not enforce cointegration, as an intermediate case between the assumption of cointegration enforced by Mark and the assumption of independence that produced Tables 3 and 4:

$$s_t = a_0 + \sum_{j=1}^{P} a_{1j} s_{t-j} + \sum_{j=1}^{P} a_{2j} f_{t-j} + \varepsilon_{1,t}, \tag{19}$$

$$f_t = b_0 + \sum_{j=1}^{P} b_{1j} s_{t-j} + \sum_{j=1}^{P} b_{2j} f_{t-j} + \varepsilon_{2,t}. \tag{20}$$

The results from the unrestricted VAR—shown in Table 5—show very little evidence of predictability except for the CHF. BG noted that for the DEM, JPY, and CAD the p values for the OUT/RW statistics are smaller at shorter horizons than they are at longer horizons, indicating less evidence of predictability at longer horizons, in contrast to Mark's basic conclusion.[17]

Kilian's (1999) primary focus was the study of the *power function* of forecasting regressions at short horizons versus long horizons rather than foreign exchange predictability per se. Such a study of power requires a null DGP. Kilian (1999) carefully mapped the monetary model to a constrained vector error correction model (VECM), which he estimated by feasible generalized least squares (FGLS) to construct bootstrapping distributions for the forecasting statistics:

$$\Delta s_t = v_e + u_{1,t} \tag{21}$$

and

$$\Delta f_t = v_f - h_2\left[s_{t-1} - f_{t-1}\right] + \sum_{j=1}^{p-1} \xi_j^{21} \Delta s_{t-j} + \sum_{j=1}^{p-1} \xi_j^{22} \Delta f_{t-j} + u_{2,t}, \tag{22}$$

[14] Kilian (1999) emphasized the importance of a drift in the out-of-sample benchmark, as discussed below.

[15] Tables 3, 4, and 5 show only a subset of the test statistics.

[16] The overturned results were from $t(A)$ and DM(A, 20) statistics, some of which Table 4 does not show.

[17] The unrestricted VAR does permit predictability, so the p values in Table 5 are the probabilities of obtaining test statistics at least as extreme as actually found, conditional on exchange rates and fundamentals following the estimated VAR.

Table 3

Results of a Forecasting Exercise Using the DGP in Equations (17) and (18) (1973:Q2–1991:Q4)

Country	Horizon	Beta	t(20)	p Value	R^2	OUT/RW	p Value	DM(20)	p Value
Canada	1	0.040	3.051	**0.095**	0.059	0.998	0.405	0.061	**0.441**
	4	0.155	2.398	0.217	0.179	1.119	0.412	–1.270	0.849
	8	0.349	2.539	0.225	0.351	1.145	0.712	–1.036	0.958
	12	0.438	1.961	0.352	0.336	1.436	0.317	–1.916	0.592
	16	0.450	1.542	0.458	0.254	1.699	0.196	–2.596	0.466
Germany	1	0.035	1.836	0.510	0.015	1.015	0.969	–0.932	0.724
	4	0.205	2.902	0.354	0.104	1.037	0.914	–1.345	0.522
	8	0.554	3.487	0.354	0.265	1.002	0.809	–0.027	0.814
	12	0.966	6.329	0.165	0.527	0.796	0.406	4.246	**0.093**
	16	1.324	9.256	**0.096**	0.762	0.524	0.113	8.719	**0.030**
Japan	1	0.047	1.396	0.516	0.020	0.988	0.477	1.571	0.286
	4	0.263	2.254	0.353	0.125	0.928	0.396	2.302	0.215
	8	0.575	3.516	0.228	0.301	0.819	0.304	3.096	0.172
	12	0.945	4.889	0.166	0.532	0.712	0.233	3.319	0.173
	16	1.273	4.919	0.216	0.694	0.574	0.142	5.126	0.109
Switzerland	1	0.074	2.681	0.210	0.051	0.997	0.642	0.066	0.704
	4	0.285	3.248	0.181	0.180	0.981	0.596	0.218	0.676
	8	0.568	4.770	**0.095**	0.336	0.917	0.458	0.703	0.621
	12	0.837	8.013	**0.032**	0.538	0.738	0.214	2.933	0.203
	16	1.086	17.410	**0.006**	0.771	0.411	**0.026**	9.650	***0.019***

NOTE: This is excerpted from Table 4 in BG. Errors for the null DGP in equations (17) and (18) were drawn with a parametric bootstrap. The out-of-sample benchmark for comparison is a random walk with drift. See the notes to Table 1 for column headings. Boldface p values denote significance at the 10 percent level. Shaded boxes indicate results that would be significant with the cointegration assumption but are no longer significant without it. Italicized p value indicates figure that was not significant with the cointegration assumption, but is now significant without it.

where the system requires $h_2 < 0$ for stability. Kilian notes that while the DGP is asymptotically equivalent to Mark's (1995) approximation, (21) and (22) will generate a different small sample distribution because of the different lag structure and estimation procedure. The results will also be sensitive to whether the estimated coefficients for the null DGP have been corrected for the finite-sample bias (see the boxed insert) caused by the persistence of the regressors, which was not compensated for in Mark's procedure.

Kilian (1999) also emphasized the importance of the treatment of the drift term in the forecasting procedures. Specifically, Mark's inconsistency in permitting a drift in the bootstrap DGP but not in the benchmark forecast biased the bootstrap critical values. In addition, comparing a fundamental forecast that has drift with a driftless random walk will conflate the contributions of the fundamentals and the drift. That is, if the out-of-sample statistics for the fundamental model are superior to those of the driftless random walk, one cannot be certain whether it is due to the contribution of the fundamentals or just the drift in the monetary model. To isolate the marginal contribution of the fundamentals, one has to allow for a drift in the benchmark forecast.

To illustrate the importance of the treatment of the drift term, Table 6—excerpted from the December 1997 working paper version of Kilian (1999)—

Table 4

Results of a Forecasting Exercise Using the DGP in Equations (17) and (18) (extended sample, 1973:Q2–1994:Q4)

Country	Horizon	Beta	t(20)	p Value	R^2	OUT/RW	p Value	DM(20)	p Value
Canada	1	0.035	3.013	0.136	0.041	0.994	0.428	0.169	0.540
	4	0.147	2.475	0.295	0.155	1.040	0.853	–0.270	0.724
	8	0.336	2.489	0.325	0.331	1.078	0.818	–0.316	0.756
	12	0.430	1.799	0.505	0.334	1.280	0.416	–0.842	0.967
	16	0.441	1.350	0.606	0.236	1.542	0.260	–1.580	0.782
Germany	1	0.038	2.269	0.431	0.021	0.998	0.715	0.117	0.713
	4	0.156	2.369	0.487	0.082	1.005	0.821	–0.124	0.836
	8	0.396	2.617	0.523	0.216	1.055	0.987	–0.286	0.930
	12	0.697	3.250	0.474	0.393	1.133	0.886	–0.340	0.961
	16	1.019	3.956	0.432	0.601	1.235	0.720	–0.518	0.967
Japan	1	0.032	1.079	0.745	0.012	0.976	0.242	1.818	0.267
	4	0.174	1.360	0.723	0.065	0.942	0.429	0.991	0.490
	8	0.422	2.093	0.609	0.182	0.895	0.485	0.986	0.587
	12	0.719	3.027	0.485	0.364	0.932	0.715	0.338	0.820
	16	0.907	3.050	0.533	0.486	1.067	0.855	–0.235	0.990
Switzerland	1	0.080	2.559	0.304	0.052	0.949	***0.035***	2.195	0.180
	4	0.287	3.195	0.250	0.175	0.838	***0.063***	1.629	0.305
	8	0.566	4.635	0.165	0.332	0.722	***0.076***	1.222	0.475
	12	0.848	7.942	**0.066**	0.538	0.455	***0.015***	1.428	0.505
	16	1.081	18.820	**0.007**	0.769	0.347	**0.007**	1.614	0.568

NOTE: This is excerpted from Table 5 in BG. Errors for the null DGP in equations (17) and (18) were drawn with a parametric bootstrap. The out-of-sample benchmark for comparison is a random walk with drift. See Table 1 notes for column headings. Boldface p values denote significance at the 10 percent level. Shaded boxes indicate results that were significant for the original sample (1973:Q2–1991:Q4), but are no longer significant with the extended sample. Italicized p values indicate the reverse—results that were not significant for the original sample (1973:Q2–1991:Q4), but now are significant with the extended sample.

presents the results from Mark's (1995) forecasting exercise using Kilian's DGP—equations (21) and (22)—with and without a drift in the random walk. The left-hand panel presents results from the drift-less random walk benchmark while the right-hand panel presents results from the random walk with drift. Contrasting the results, using a drift in the benchmark eliminates any evidence of predictability for the JPY case but increases the predictability in the CAD and CHF, especially at short horizons.

Indeed, both BG and Kilian (1999) took issue with the whole idea of finding predictability in long-horizon regressions. BG show that a linear model offers no more predictability at long horizons than at short horizons. Kilian (1999) focused on the question of whether long-horizon regressions truly have greater power to find predictability than short-horizon tests. In particular, he extended the analytic work done by BG with Monte Carlo experiments that showed that increasing evidence of predictability at long horizons was due to the fact that such tests were more likely to err in favor of finding predictability where there was none, rather than really being better at finding latent predictability. In econometric jargon, the results were due to *size distortions* rather than *power gains*. To summarize: Both BG and Kilian (1999) conclude that it doesn't help to increase the forecast horizon if the DGP is linear.

Table 5

Results of a Forecasting Exercise Using the BG DGP No. 2, an Unrestricted VAR, Equations (19) and (20) (1973:Q2–1994:Q4)

Country	Horizon	Beta	t(20)	p Value	R^2	OUT/RW	p Value	DM(20)	p Value
Canada	1	0.035	3.013	0.426	0.041	0.994	0.557	0.169	0.597
	4	0.147	2.475	0.873	0.155	1.040	0.832	–0.270	0.799
	8	0.336	2.489	0.957	0.331	1.078	0.845	–0.316	0.827
	12	0.431	1.799	0.989	0.334	1.284	0.907	–0.842	0.877
	16	0.440	1.352	0.988	0.236	1.542	0.946	–1.580	0.914
Germany	1	0.038	2.269	0.740	0.021	0.998	0.497	0.117	0.483
	4	0.156	2.369	0.765	0.082	1.005	0.534	–0.124	0.547
	8	0.396	2.617	0.759	0.216	1.055	0.590	–0.286	0.570
	12	0.697	3.255	0.691	0.393	1.133	0.661	–0.340	0.585
	16	1.019	3.956	0.614	0.601	1.235	0.730	–0.518	0.636
Japan	1	0.032	1.079	0.986	0.012	0.976	0.359	1.818	0.214
	4	0.174	1.365	0.982	0.065	0.942	0.504	0.991	0.433
	8	0.422	2.093	0.938	0.182	0.895	0.512	0.986	0.475
	12	0.719	3.027	0.855	0.364	0.932	0.583	0.338	0.604
	16	0.907	3.050	0.843	0.486	1.067	0.714	–0.235	0.692
Switzerland	1	0.080	2.559	0.513	0.052	0.949	**0.062**	2.195	0.150
	4	0.287	3.195	0.457	0.175	0.838	0.103	1.629	0.260
	8	0.566	4.635	0.300	0.332	0.722	**0.085**	1.222	0.330
	12	0.848	7.942	**0.090**	0.538	0.455	**0.010**	1.428	0.301
	16	1.081	18.820	**0.001**	0.769	0.347	**0.004**	1.614	0.300

NOTE: This is excerpted from Table 6 in BG. Errors for the null DGP in equations (19) and (20) were drawn with a parametric bootstrap. The out-of-sample benchmark for comparison is a random walk with drift. See Table 1 notes for column headings. Boldface p values denote significance at the 10 percent level. Shaded cell indicates result that was significant in Table 4, but is no longer significant with the unrestricted VAR as the DGP.

The work of BG and Kilian (1999) shows that the results of the forecasting exercise were sensitive to a number of factors, including the data sample, the assumption of cointegration in the DGP, the lag structure of the DGP, the benchmark for out-of-sample comparison, and whether one corrects the DGP for bias generated by persistent regressors.[18] Indeed, their conclusions on predictability are very similar. BG conclude that failure to impose cointegration leaves only weak evidence of predictability, and that is at predominantly short horizons. Kilian (1999) similarly concludes that with properly generated critical values, there is some evidence that monetary fundamentals predict foreign exchange rates but no evidence of more forecastability at longer horizons.

Panel Studies

When alternative explanations—i.e., predictability or no predictability—seem to fit the data equally well, employing additional data often illuminates the issue. In the present case, one might combine evidence from many exchange rates in a panel study of predictability, under the assumption that the exchange rates are either predictable from fundamentals for all the countries or predictable for none of them. Groen (2000) and Mark and Sul (2001) aggregated information about the predictability of exchange rates across countries. Rapach and Wohar

[18] Groen (1999) also reports the fragility of Mark's (1995) results to the chosen sample.

Table 6

Results of a Forecasting Exercise Using Kilian's DGP, Equations (21) and (22) (1973:Q2–1994:Q4)

		Benchmark: driftless RW p values			Benchmark: RW with drift p values		
Country	Horizon	t(20)	OUT/RW	DM(20)	t(20)	OUT/RW	DM(20)
Canada	1	**0.054**	**0.045**	**0.065**	**0.054**	**0.048**	**0.055**
	4	0.132	0.147	0.139	0.131	**0.055**	**0.050**
	8	0.172	0.139	0.145	0.172	**0.052**	**0.022**
	12	0.286	0.341	0.247	0.287	0.129	0.116
	16	0.392	0.539	0.469	0.391	0.250	0.345
Germany	1	0.420	0.348	0.470	0.346	0.352	0.537
	4	0.486	0.329	0.675	0.395	0.309	0.407
	8	0.520	0.334	0.316	0.402	0.308	0.303
	12	0.473	0.268	0.249	0.359	0.272	0.274
	16	0.428	0.206	0.231	0.344	0.281	0.299
Japan	1	0.771	**0.080**	**0.042**	0.629	0.269	0.360
	4	0.738	**0.079**	**0.071**	0.589	0.276	0.290
	8	0.620	**0.072**	**0.083**	0.473	0.217	0.221
	12	0.497	**0.099**	0.142	0.362	0.276	0.277
	16	0.557	0.112	0.173	0.403	0.389	0.346
Switzerland	1	0.150	0.107	0.123	0.150	**0.061**	**0.081**
	4	0.132	0.117	0.130	0.132	**0.091**	0.102
	8	**0.095**	0.113	0.137	**0.095**	**0.089**	0.112
	12	**0.036**	**0.046**	**0.092**	**0.035**	**0.032**	**0.064**
	16	**0.001**	**0.003**	**0.073**	**0.001**	**0.007**	**0.049**

NOTE: This table is excerpted from Tables 2 and 3 in the December 1997 working paper version of Kilian (1999). The left-hand panel uses a driftless random walk for the DGP and out-of-sample benchmark. The right-hand panel uses a random walk with drift for the DGP and out-of-sample benchmark. The columns display the p values for the t statistic with a 20-period window, ratio of RMSEs, and Diebold-Mariano statistics for differences in RMSEs, respectively. Boldface p values denote significance at the 10 percent level.

(2001b) have examined whether such aggregation is appropriate.

Groen (2000) examines the question of whether exchange rates are cointegrated with fundamentals using both rate-by-rate Johansen (1991) cointegration tests and the Levin and Lin (1993) panel unit root tests on 4 subsets of 14 exchange rates against the USD: (i) all 14; (ii) the G-10; (iii) the G-7; and (iv) the European Monetary System (EMS). The rate-by-rate Johansen tests reject the null of no cointegration in about one-third of the cases using either the USD or the DEM as numeraire currency; this suggests that "cointegration isn't widespread." The more powerful Levin-Lin (1993) panel test, however, rejects the null of no cointegration jointly for all the rates at the 5 percent level, using either the USD or DEM as numeraire for the 14-country panel.[19] While the results for smaller panels are often insignificant, the overall conclusion is that the most powerful tests are supportive of cointegration.

The growing consensus against long-horizon prediction regressions and the econometric complications caused by overlapping forecast errors led Mark and Sul (2001) to eschew the search for long-run predictability in favor of a one-period-ahead panel regression of quarterly data on 18 exchange

[19] Taylor, Peel, and Sarno (2001) note that panel unit root tests tend to reject the null of a unit root if even one of the series is stationary, because the null hypothesis is that all of the series have a unit root.

rates and fundamentals. The sample started in 1973:Q1 and ended in 1997:Q1. Mark and Sul (2001) first test for and find evidence of cointegration with a panel dynamic OLS framework, which controls for asymptotic bias in the forecast statistics. This cointegration finding is used to construct the bootstrapped data that corrects coefficients for persistent regressor bias and evaluates the statistical significance of Theil U statistics. The forecasting equation is the multi-exchange rate analogue to those used in previous papers, a one-period-ahead panel regression, estimated by seemingly unrelated regressions (SUR) over expanding samples.[20] The system could be written as follows:

$$s_{i,t+1} - s_{i,t} = \beta\left(f_{i,t} - s_{i,t}\right) + \varepsilon_{i,t+1} \tag{23}$$

$$\varepsilon_{i,t+1} = \gamma_i + \theta_{t+1} + u_{i,t+1},$$

where $s_{i,t}$ and $f_{i,t}$ are the log exchange rate and the log fundamentals of exchange rate i at time t, γ_i is an exchange-rate-specific error, θ_t is a time-specific error, and $u_{i,t}$ is an idiosyncratic error. Three exchange rates are considered as numeraire for the system: the USD, the CHF, and the JPY. In the one-step-ahead forecasting exercises, Mark and Sul (2001) find that monetary fundamentals have a small but statistically significant amount of predictability—using Theil U statistics—when the USD or CHF is numeraire but none when the JPY is the standard. They also find that monetary fundamentals predict somewhat better than PPP fundamentals.

Both the Groen (2000) and Mark and Sul (2001) studies pooled data across countries to try to bring more power to answering the question of how well monetary models predict the exchange rate. The practice of pooling data across countries assumes, of course, that the same DGP produces the data for all the countries. Such assumptions are called "homogeneity assumptions." If the DGP is different across countries, however, then pooling the data can lead to incorrect inference. Using the Mark and Sul (2001) data set, Rapach and Wohar (2001b) first confirm previous results that the monetary model fits very poorly in country-by-country estimation during the floating rate period (1973:Q1–1997:Q1). "It is difficult to overstate how poorly the monetary model performs…on a country by country basis" (Rapach and Wohar, 2001b, p. 3). In contrast, however, pooled estimates do support the monetary model, as in Mark and Sul (2001). Next, the authors formally test whether the cross-country homogeneity assumptions are justified. That is, is it likely that one DGP could have produced the disparate coefficient estimates from the 14 different exchange rates? A Wald test rejects this one-DGP hypothesis for most subsets of countries (Mark, Ogaki, and Sul, 2000). And a Monte Carlo study shows that it is very plausible that heterogeneous DGPs—fit to the 14 exchange rate/fundamental processes—could produce pooled parameter estimates similar to those found in the real data. These findings cast doubt on the wisdom of pooling data across countries and the reliability of the conclusions.

Bolstering the argument in favor of pooling, however, is the fact that the pooled parameter estimates are as good as the country-by-country forecasts at short horizons and better at long horizons. Rapach and Wohar (2001b) cite Pesaran, Shin, and Smith (1999) as arguing that omitted variables and measurement error might lead to the false rejection of homogeneity restrictions and that pooling might still be appropriate and helpful under such circumstances. Ultimately, Rapach and Wohar (2001b) conclude that researchers could reasonably differ about the fit of the monetary model of exchange rates during the post-Bretton Woods period.

Long Spans of Data

Combining evidence from many countries in a panel study is one way to increase the available data to determine whether exchange rates are cointegrated with fundamentals. Another approach is to simply use much longer spans of data. Rapach and Wohar (2001a) took this latter approach, using exchange rate, money, and output data from 14 industrialized countries, over a span as long as 115 years (1880-1995), to investigate the long-run relationship between these variables. Table 7 summarizes the results from Rapach and Wohar (2001a).

First, Rapach and Wohar (2001a) noted that if exchange rates are to be predicted from relative money and output, some combination of the three variables $\{s_t, (m_t - m_t^*), (y_t - y_t^*)\}$ must be stationary (I(0)). If no combination is I(0), then the error from any forecast will become arbitrarily big as time goes on, bigger than the benchmark error. If one of these variables $\{s_t, (m_t - m_t^*), (y_t - y_t^*)\}$ is I(1), for example, while the other two are I(0), then no linear combination can be I(0) and the monetary model can be

[20] In expanding samples, one period is added to the in-sample data used to estimate coefficients before each forecast to give the model the maximum amount of data with which to construct out-of-sample forecasts.

Table 7

Summary of Rapach and Wohar (2001a)

Country	Sample period	Unit root tests on $\{s_t, m_t - m_t^*, y_t - y_t^*\}$	DOLS cointegrating vector	Tests that support cointegration	Unit root tests on $s_t - [m_t - m_t^*] + [y_t - y_t^*]$	$(f_t - s_t)$ Predicts Δs_{t+1} in a VECM	$(f_t - s_t)$ Predicts Δf_{t+1} in a VECM	Four-quarter-ahead Theil U statistics
Australia	1880-1995		{0.45, –0.19}					
Belgium	1880-1989		{1.00, –1.01}	1 of 4	Moderate support	Yes	No	0.98
Canada	1880-1995		{0.13, 0.10}					
Denmark	1885-1995	Reject the model						
Finland	1911-1995		{1.01, NA}	3 of 4	Moderate support	Yes	No	1.02
France (trend)	1880-1989		{1.03, –1.16}	4 of 4	Strong support	Yes	Yes	1.02
Italy	1880-1995		{0.96, –1.34}	4 of 4	Strong support	Yes	No	0.94
Netherlands	1900-1992		NA	NA	Strong support			
Norway	1899-1995	Reject the model						
Portugal	1929-1995		{1.07, NA}	3 of 4	Moderate support	No	Yes	1.01
Spain	1901-1995		{0.86, –1.29}	4 of 4	Strong support	No	Yes	1.03
Sweden	1880-1995	Reject the model						
Switzerland (trend)	1880-1995		{0.86, –1.30}	1 of 4	Weak support	Yes	Yes	0.99
United Kingdom	1880-1995		{0.45, –0.99}					

NOTE: This table summarizes the results from Rapach and Wohar (2001a). "Trend" after a country name indicates that a trend was permitted in cointegrating relations. Column 3 displays "Reject" if univariate unit root tests on the 3 series $\{s_t, m_t - m_t^*, y_t - y_t^*\}$ permit one to reject that a linear combination of them can be stationary. Rejecting this hypothesis permits one to immediately reject the monetary model of exchange rates. Column 4 shows dynamic OLS estimates of the cointegrating vector (Stock and Watson, 1993). The monetary model implies a vector of {1, –1}. NA indicates that either all variables were stationary (the Netherlands) or that relative output $(y_t - y_t^*)$ is stationary (Finland and Portugal). Column 5 provides the number of cointegration tests (out of 4) that support the hypothesis that exchange rates are cointegrated with fundamentals. Column 6 shows the degree to which tests on $s_t - [m_t - m_t^*] + [y_t - y_t^*]$ reject the null of a unit root and thereby support the monetary model. Columns 7 and 8 show whether the coefficients λ_1 and λ_2 from the VECM described by equations (25) and (26) are statistically significant, respectively. Column 9 provides the one-year-ahead out-of-sample Theil U statistic in each case. Statistics less than 1 indicate that the monetary model outforecast the random walk benchmark.

rejected. Such unit root tests alone enable the authors to reject the monetary model for Denmark, Norway, and Sweden (Table 7, column 3).

Second, Rapach and Wohar (2001a) go on to estimate cointegrating vectors—using four different methods—for the following equation:

$$(24) \qquad s_t = \beta_0 + \beta_1\left(m_t - m_t^*\right) + \beta_2\left(y_t - y_t^*\right) + \varepsilon_t ;$$

they test—using four different cointegration tests—whether those vectors are consistent with the vector implied by the simple monetary model $\{\beta_1 = 1, \beta_2 = -1\}$ (Table 7, columns 4 and 5).[21] All four tests find evidence of cointegration for France, Spain, and Italy; three tests find evidence for Finland and Portugal; and one test finds evidence for Belgium and Switzerland. The estimated coefficients are often close to those implied by the simple monetary model $\{\beta_1 = 1, \beta_2 = -1\}$.

Third, the authors use unit root tests on the residuals from the error correction term $\{s_t - (m_t - m_t^*) + (y_t - y_t^*)\}$, which is implied by the simple monetary model. A rejection of the unit root hypothesis is interpreted as supporting the "simple," "long-run" monetary model. The tests produce strong support for the monetary model for the Netherlands, France, Italy, and Spain; moderate support for Belgium, Finland, and Portugal; and weak support for Switzerland (Table 7, column 6). The authors caution, however, that deviations from monetary fundamentals can be substantial and very persistent.

Fourth, a VECM is estimated to investigate the dynamics of the relation between exchange rates and the fundamentals. The VECM can be written as follows:

(25)

$$\Delta s_t = \lambda_1\left(f_{t-1} - s_{t-1}\right) + \sum_{i=1}^{p} a_i \Delta s_{t-i} + \sum_{i=1}^{p} b_i \Delta f_{t-i} + \mu_1 + \varepsilon_{1,t}$$

(26)

$$\Delta f_t = \lambda_2\left(f_{t-1} - s_{t-1}\right) + \sum_{i=1}^{p} c_i \Delta s_{t-i} + \sum_{i=1}^{p} d_i \Delta f_{t-i} + \mu_2 + \varepsilon_{2,t}.$$

Note that fundamentals predict exchange rates in the expected way if either

$$\lambda_1 > 0 \text{ or } \sum_{i=1}^{P} b_i > 0.$$

Similarly, exchange rates predict fundamentals in the expected way if either

$$\lambda_2 < 0 \text{ or } \sum_{i=1}^{P} c_i > 0.$$

Rapach and Wohar (2001a) find that the error correction term ($f_{t-1} - s_{t-1}$) predicts exchange rates for Belgium, Finland, and Italy ($\lambda_1 > 0$, $\lambda_2 = 0$). In VECM jargon, monetary fundamentals are said to be *weakly exogenous* for these cases. For Portugal and Spain, the exchange rate is weakly exogenous, the error correction term predicts future fundamental changes but not exchange rate changes ($\lambda_1 = 0$, $\lambda_2 < 0$). The error correction term predicts both future fundamentals and exchange rate changes for France and Switzerland ($\lambda_1 > 0$, $\lambda_2 < 0$). These results are summarized in columns 7 and 8 of Table 7.

Finally, the authors pursue the usual out-of-sample forecasting exercises and find that their Theil U statistics, DM statistics, and *encompassing regressions* show evidence of predictability—beyond the random walk with drift—for Belgium, Italy, and Switzerland and some evidence for Finland (from encompassing regressions not shown in Table 7).[22] Not surprisingly, cases in which the VECM showed the exchange rate to be weakly exogenous are out-of-sample forecasting failures. The authors note, however, that forecasting failure can still be consistent with the long-run monetary model if deviations from fundamentals predict future fundamentals (Table 7, column 9).

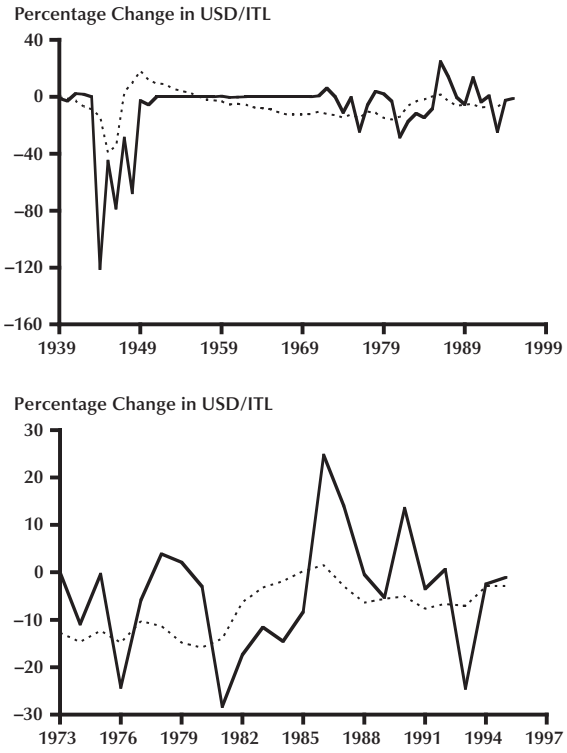
Figure 1 uses the case in which the monetary model has the best out-of-sample fit—the case of the USD/Italian lira (ITL)—to illustrate how little of the variation in the one-year-ahead exchange rate change the monetary model predicts. The top panel shows exchange rate changes and recursive, out-of-sample forecasts from 1939 to the end of the sample in 1995. The bottom panel shows the same data from the beginning of the floating exchange rate era, 1973-95. The figure appears to show that the monetary model forecasts best in extreme circumstances, such as those seen during the high inflation

[21] Rapach and Wohar (2001b) defined exchange rates to be the foreign currency price of domestic currency—the inverse of the definition used previously in this paper—and defined the fundamentals as the negative of Mark's fundamentals. For consistency, this paper will use Mark's definitions of the exchange rate and fundamentals. Also, note that Rapach and Wohar (2001b) imposed $\beta_2 = 0$ in their estimate of the cointegrating vector for Finland and Portugal because relative output was found to be I(0) in those cases.

[22] An encompassing regression evaluates the predictive content of forecasting techniques by testing whether realized values of the exchange rate depend positively on predicted values from one or more forecasting techniques.

Figure 1

Forecasts of the Change in the USD/ITL Exchange Rate



NOTE: The upper panel depicts the percentage annual change in the USD/ITL exchange rate (solid line) and the predicted value (dashed line) from the monetary model from 1939 to 1995. The lower panel depicts the same data over the floating rate period, 1973-95.

that Italy suffered during World War II or during 1986-87, when the dollar weakened again after a period of unusual strength. In almost all periods, however, the monetary model explains very little of the variation in one-year-ahead exchange rate changes.

Data Revisions

Faust, Rogers, and Wright (2001) examine the impact of yet another issue: data revisions. Previous studies had all assumed that the data they used would be fixed and known to a forecaster. In fact, macroeconomic data such as money supply and output figures are often extensively revised, meaning that data depends on the date on which the series were obtained. In other words, if one obtained U.S. output data in April 1992—as the authors infer that Mark did—one might have a different estimate of U.S. output growth for 1991:Q1 than if one obtained data in April 1993. To investigate the effect of data revisions on forecasting exercises with monetary models, Faust, Rogers, and Wright (2001) obtained 38 data sets, representing the best estimates of the data as it stood on different dates from April 1988 to October 2000.

With these 38 data sets, they first attempted to see whether one would obtain the same inference as Mark (1995) by holding back the final 40 quarters of data from each set for an out-of-sample forecasting exercise. That is, each of the 38 data sets were different from each other both because of data revisions and the fact that the final 40 quarters of data would be over different periods. They found that the only vintages of the data that would have produced significant long-horizon predictability were those in a 2-year window around April 1992, the time that Mark collected his data set. The decrease in predictability for other data sets was due to both sample periods—as noted by Kilian (1999)—and data revisions.

To isolate the marginal effect of data revisions, the authors fixed the sample period and compared results using more- and less-revised data. The more-revised data showed less evidence of predictability. Theil U statistics rise and p values fall as the data are revised. For example, data revisions made the 16-quarter Theil U statistic rise from 0.52 to 0.64 in the German case and from 0.55 to 0.69 in the Japanese case. And the p values for these cases rise above 0.1. Similarly, the authors estimate a portfolio-balance model using inflation, interest rates, and cumulated trade balances—as in Chinn and Meese (1995)—and they find that data revisions have even larger effects than in the monetary model.

Next Faust, Rogers, and Wright (2001) investigate the quality of "real-time" data forecasts. Real-time forecasts use the latest revision of data available at any given point in time to estimate the parameters of the model and make forecasts. That is, real-time exercises can use a different set of data for each forecast period. In contrast, the other forecasting exercises (e.g., Mark, 1995) used one data set—the latest revisions available when the research is done—to estimate the equation and make forecasts. Faust, Rogers, and Wright find that real-time data provide better out-of-sample predictive power—according to out-of-sample relative RMSEs—in almost every case than the latest data revision. Perhaps this should not surprise us. If the exchange rate changes do

depend on market participants' expectations of future monetary fundamentals (i.e., equation (14)), which are based on currently available (real-time) data, then real-time data should provide better estimates of market expectations.

Finally, the authors find that Federal Reserve forecasts of future variables sometimes outperform actual future values of those independent variables in an MR-type regression multi-period forecast. This is ironic. MR sought to give the monetary models the best possible chance to forecast well by replacing forecasts of future independent variables with actual values. But, (at least some) forecasts of fundamentals predict exchange rates better than the future values of those variables.

The study concludes that both the particular sample period used by Mark (1995) and the particular vintage of data revisions combined to produce better out-of-sample forecasting performance than most data samples and/or data revisions would have. Indeed, only data sets constructed in about 1992 would have shown evidence of long-horizon predictability. Faust, Rogers, and Wright (2001) speculate that evidence of forecastability is actually an artifact of data mining, the tendency to test multiple models on one set of data until, by chance, positive results are found. Finally, for a given fixed sample, real-time data would have produced better forecasts than the latest data revisions. Hindsight turns out to handicap the forecasts rather than to improve them.

Nonlinear Models

The monetary model is intuitively appealing but clearly explains very little exchange rate variability. One explanation for the weak relation is that exchange rates are relatively insensitive to monetary fundamentals close to equilibrium values but tend to strongly revert to those fundamentals when the deviation is large. Taylor and Peel (2000), Taylor, Peel, and Sarno (2001), and Kilian and Taylor (2001) investigate the plausibility of this characterization with nonlinear models.

Taylor and Peel (2000) estimate a nonlinear model of quarterly exchange rates and monetary fundamentals for the British pound (GBD)/USD and DEM/USD exchange rates from 1973:Q1 to 1996:Q4. They find that the exponential smooth transition autoregressive (ESTAR) model (Granger and Teräsvirta, 1993) parsimoniously describes the deviation of the exchange rate from monetary fundamentals. This model predicts that the exchange rate change will be nearly unpredictable when the deviation from fundamentals is small, but will strongly revert toward those fundamentals when the deviation is big. The authors use this to characterize the degree of over- and undervaluation of the exchange rates during the modern period of floating exchange rates. Similarly, Taylor, Peel, and Sarno (2001) show that the same model fits real exchange rates well and explains deviations from PPP.

Kilian and Taylor (2001) note that a convincing explanation for the nonlinear dynamics of the ESTAR model is lacking. The authors suggest a candidate model in which uncertainty about the fundamental value of the exchange rate deters agents from speculating against small deviations from fundamentals.[23] Monte Carlo studies show that there is more predictability for plausible DGPs at the one- and two-year horizons, so long-horizon tests are useful in such an environment. Further, if the ESTAR model is the true DGP, then all past tests of long-horizon predictability are invalid because they assume a linear null DGP, which is incorrect. Consistent with this prediction, the authors find that in-sample evidence of predictability from seven OECD countries increases "dramatically" with the forecast horizon.[24] Yet, the authors are still unable to find evidence of out-of-sample predictability. They ascribe this to the low power of out-of-sample tests, given the short span of post-Bretton Woods data and the rarity of large departures from fundamentals during that time.

WHY DOESN'T THE MONETARY MODEL PREDICT WELL?

One obvious problem is that three of the building blocks of the monetary model, money demand equations, purchasing power parity (PPP), and uncovered interest parity (UIP) do not work very well (Engel, 1996 and 2000). Money demand equations have proven unstable, especially in the United States (Friedman and Kuttner, 1992), but changing the numeraire currency doesn't seem to help the monetary model much.

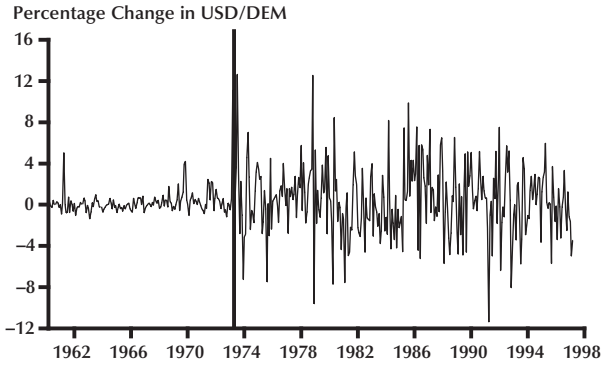
But that begs the question as to why PPP and UIP perform so poorly. Why are floating exchange

[23] Kilian and Taylor (2001) assume that the fundamental value is a function of relative prices rather than money and output, as in the monetary model.

[24] Mark and Sul (2002) find that long-horizon regressions can have asymptotic power advantages over one-period-ahead procedures in cases similar to those found in foreign exchange forecasting. Their Monte Carlo experiments show that the phenomenon might be even more common in finite samples.

Figure 2

Monthly Percentage Changes in the Real USD/DEM Exchange Rate



NOTE: These changes become much more volatile after March 1973, the end of the Bretton Woods system of fixed exchange rates. The vertical line denotes this break date in the series.

rates so volatile and unrelated to prices and interest differentials? Many researchers have claimed that volatile expectations or departures from rationality are likely to account for the failure of exchange rate models. For example, Frankel (1996) argues that exchange rates are detached from fundamentals by swings in expectations about future values of the exchange rate. These fluctuations in exchange rates are essentially bubbles, of the type discussed in the second section of the paper. Four pieces of evidence suggest that expectations are to blame for such behavior: (i) Survey measures of exchange rate expectations are very poor forecasters and the expectations themselves are frequently not internally consistent (Frankel and Froot, 1987; Sarno and Taylor, 2001); (ii) Failure of expectations to be rational is often blamed for the failure of UIP (Engel, 1996); (iii) Trend-following trading rules appear to make risk-adjusted excess returns, in apparent violation of the efficient markets hypothesis (Neely, 1997; Neely, Weller, and Dittmar, 1997); (iv) Switching from a fixed exchange rate to a floating rate—which changes the way expectations are formed—changes the behavior of nominal and real exchange rates and the ability of UIP to explain exchange rate changes.

This latter point requires some explanation. Fixed exchange rates anchor investor sentiment about the future value of a currency because of the government's commitment to stabilize its value. If expectations are based on fundamentals, rather than irrationally changing expectations, then the relationship between fundamentals and exchange rates should be the same under a fixed exchange rate regime as it is under a floating regime. This is not the case. Countries that move from floating exchange rates to fixed exchange rates experience a dramatic change in the relationship between prices and exchange rates. Specifically, real exchange rates (exchange rates adjusted for inflation in both countries) are much more volatile under floating exchange rate regimes, where expectations are not tied down by promises of government intervention (Mussa, 1986). Figure 2 illustrates a typical case: When the German government ceased to fix the DEM to the USD in March 1973, the variability in the real USD/DEM exchange rate increased dramatically. This result suggests that, contrary to the efficient markets hypothesis, swings in investor expectations may detach exchange rates from fundamental values in the short run. Similarly, UIP seems to do such a poor job explaining USD exchange rates while doing a pretty good job with semi-fixed rates such as those found in the EMS (Flood and Rose, 1996). Indeed, Flood and Rose (1999) develop a UIP-based model of the exchange rate that explains why UIP—and exchange rate forecasts—might perform poorly in the short term even with perfectly rational agents. UIP also performs better over long horizons than over short horizons (Meredith and Chinn, 1998; Alexius, 2001). The common thread among these cases is that fluctuations in short-term expectations do not affect the model's performance.

CONCLUSION: THE BIG PICTURE

The seminal work of MR showed that monetary models were unable to forecast exchange rates better than a no-change forecast. Since then, a small army of researchers has attempted to forecast exchange rates with the analytically attractive monetary model. Initial attempts were strikingly unsuccessful; Mark (1995), however, appeared to show that monetary fundamentals could predict exchange rate changes at three- to four-year horizons. Kilian (1999), Berkowitz and Giorgianni (2001), and Faust, Rogers, and Wright (2001) subsequently criticized the underlying assumptions of Mark's study with respect to the stationarity of the data, robustness to sample period, appropriate benchmark for comparison, and the vintage of the data. Attempts to forecast with panel studies (Mark and Sul, 2001; Rapach and Wohar, 2001b) or very long samples (Rapach and Wohar, 2001a) have failed to establish the existence of predictability beyond reasonable doubt.

Other research suggests that exchange rates might be nonlinearly mean reverting to fundamentals and that the intuitively appealing monetary model might therefore provide better predictions when exchange rates deviate substantially from fundamentals (Kilian and Taylor, 2001). Such models also imply that long-horizon regressions might be more informative than short-horizon regressions.

How should one interpret these disparate results? The monetary model is intuitively appealing, and monetary variables likely influence exchange rate changes. But, like all models, it simplifies reality. The literature on exchange rate forecasting has shown that the amount of exchange rate variation explained by monetary models is—at most—small. Further research will doubtless continue to attempt to quantify this predictability. Other future research might profitably explore the way that expectations of asset prices are formed and the factors—such as dispersion of belief, risk aversion, and transactions costs—that permit extreme nominal exchange rate variability by hindering arbitrage.

Other conclusions that one might draw depend on the purpose of these forecasting exercises, which is little discussed in the literature. Researchers have usually motivated their work by the desire to evaluate monetary models of the exchange rate (MR) and have cautioned that they are not trying to build the best possible forecasting model (Mark and Sul, 2001); but the negative results for monetary models have nonetheless produced a conventional wisdom in the profession that exchange rate changes cannot be forecast—or cannot be forecast using macroeconomic fundamentals. It is not clear, however, that this is true.

It might seem obvious that both policymakers and firms would want to forecast exchange rates, but it isn't entirely clear why they would wish to do so with monetary fundamentals. Policymakers might wish to forecast exchange rates because of their influence on variables of more direct interest such as output and inflation. But why not directly forecast output and inflation if that is the case? Or, firms might wish to forecast exchange rates to make asset allocation decisions. But that would require a forecast of deviations from UIP, not of exchange rates themselves. Surely the reason for exchange rate forecasts will influence the method of evaluation and the value of those projections. Future research should address this topic.

REFERENCES

Alexius, Annika. "Uncovered Interest Parity Revisited." *Review of International Economics*, August 2001, *9*(3), pp. 505-17.

Berben, Robert-Paul and van Dijk, Dick. "Does the Absence of Cointegration Explain the Typical Findings in Long Horizon Regressions?" Unpublished manuscript, Tinbergen Institute, 6 April 1998.

Berkowitz, Jeremy and Giorgianni, Lorenzo. "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, February 2001, *83*(1), pp. 81-91.

_____ and Kilian, Lutz. "Recent Developments in Bootstrapping Time Series." *Econometric Reviews*, February 2000, *19*(1), pp. 1-48.

Bilson, John F.O. "Rational Expectations and the Exchange Rate," in Jacob A. Frenkel and Henry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*. Reading, MA: Addison-Wesley Press, 1978.

Blough, Stephen R. "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples." *Journal of Applied Econometrics*, July-September 1992, *7*(3), pp. 295-308.

Canova, Fabio. "Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model." *Journal of Economic Dynamics and Control*, January-March 1993, *17*(1/2), pp. 233-61.

Cheung, Yin-Wong; Chinn, Menzie D. and Pascual, Antonio Garcia. "Empirical Exchange Rate Models of the Nineties: Are Any Fit To Survive?" Working paper, University of California, Santa Cruz, 28 May 2002.

Chinn, Menzie D. and Meese, Richard A. "Banking on Currency Forecasts: How Predictable Is Change in Money?" *Journal of International Economics*, February 1995, *38*(1/2), pp. 161-78.

Clarida, Richard H. and Taylor, Mark P. "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors." *Review of Economics and Statistics*, August 1997, *79*(3), pp. 353-61.

_____; Sarno, Lucio; Taylor, Mark P. and Valente, Giorgio. "The Out-of-Sample Success of Term Structure

Models as Exchange Rate Predictors: A Step Beyond." *Journal of International Economics*, 2003 (forthcoming).

Diebold, Francis X. and Mariano, Roberto S. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, July 1995, *13*(3), pp. 253-63.

Dornbusch, Rudiger. "Expectations and Exchange Rate Dynamics." *Journal of Political Economy*, December 1976, *84*(6), pp. 1161-76.

Engel, Charles. "Long-Run PPP May Not Hold After All." *Journal of International Economics*, August 2000, *51*(2), pp. 243-73.

_____ . "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence." *Journal of Empirical Finance*, June 1996, *3*(2), pp. 123-92.

Evans, M.D.D. and Lyons, R.K. "Order Flow and Exchange Rate Dynamics." Unpublished manuscript, Georgetown University, December 1999.

Fair, Ray C. "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors." *Econometrica*, May 1970, *38*(3), pp. 507-16.

_____. "Evaluating the Information Content and Money Making Ability of Forecasts from Exchange Rate Equations." Unpublished manuscript, Yale University, May 1999.

Faust, Jon. "Near Observational Equivalence and Unit Root Processes: Formal Concepts and Implications." International Finance and Economics Discussion Paper 1993-447, Board of Governors of the Federal Reserve System, 1993.

_____; Rogers, John H. and Wright, Jonathan H. "Exchange Rate Forecasting: The Errors We've Really Made." International Finance Discussion Paper 2001-714, Board of Governors of the Federal Reserve System, December 2001.

Flood, Robert P. and Rose, Andrew K. "Fixes: Of the Forward Discount Puzzle." *Review of Economics and Statistics*, November 1996, *78*(4), pp.748-52.

_____ and _____. "Understanding Exchange Rate Volatility Without the Contrivance of Macroeconomics." *Economic Journal*, November 1999, *109*(459), pp. F660-72.

Frankel, Jeffrey. "How Well Do Foreign Exchange Markets Work: Might a Tobin Tax Help?" in Mahbub ul Haq, Inge Kaul, and Isabelle Grunberg, eds., *The Tobin Tax: Coping with Financial Volatility*. Oxford: Oxford University Press, 1996.

_____ and Froot, Kenneth A. "Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations." *American Economic Review*, March 1987, *77*(1), pp. 133-53.

Frenkel, Jacob A. "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence." *Scandinavian Journal of Economics*, 1976, *78*(2), pp. 200-24.

Friedman, Benjamin M. and Kuttner, Kenneth N. "Money, Income, Prices, and Interest Rates." *American Economic Review*, June 1992, *82*(3), pp. 472-92.

Granger, Clive W.J. and Newbold, P. "Spurious Regressions in Econometrics." *Journal of Econometrics*, July 1974, *2*(2), pp. 111-20.

_____ and Teräsvirta, Timo. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press, 1993.

Groen, Jan J.J. "The Monetary Exchange Rate Model as a Long-Run Phenomenon." *Journal of International Economics*, December 2000, *52*(2), pp. 299-319.

_____. "Long-Horizon Predictability of Exchange Rates: Is It For Real?" *Empirical Economics*, August 1999, *24*(3), pp. 451-69.

Hooper, Peter and Morton, John E. "Fluctuations in the Dollar: A Model of Nominal and Real Exchange Rate Determination." *Journal of International Money and Finance*, April 1982, *1*, pp. 39-56.

Horvath, Michael T.K. and Watson, Mark W. "Testing for Cointegration When Some of the Cointegrating Vectors Are Prespecified." *Econometric Theory*, December 1995, *11*(5), pp. 984-1014.

Johansen, Soren. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, November 1991, *59*(6), pp. 1551-80.

Kendall, M.G. "Note on Bias in the Estimation of Autocorrelation." *Biometrika*, December 1954, *41*(3/4), pp. 403-04.

Kilian, Lutz. "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" Working paper, University of Michigan, December 1997. <www.econ.lsa.umich.edu/Research/wp1997.html>.

_____. "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics*, September/October 1999, *14*(5), pp. 491-510.

_____ and Taylor, Mark. "Why Is It So Difficult to Beat Random Walk Forecasts of Exchange Rates?" Unpublished manuscript, October 2001.

Lane, Philip R. "The New Open Economy Macroeconomics: A Survey." *Journal of International Economics*, August 2001, *54*(2), pp. 235-66.

Levin, Andrew and Lin, Chien-Fu. "Unit Root Tests in Panel Data: New Results." Working Paper 93-56, University of California at San Diego Department of Economics, December 1993.

Lucas, Robert E. "Interest Rates and Currency Prices in a Two-Country World." *Journal of Monetary Economics*, November 1982, *10*(3), pp. 335-59.

Mankiw, N. Gregory and Shapiro, Matthew D. "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models." *Economics Letters*, 1986, *20*, pp. 139-45.

Mark, Nelson C. *International Macroeconomics and Finance: Theory and Econometric Methods*. Malden, MA: Blackwell Publishers, 2001.

_____. "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability." *American Economic Review*, March 1995, *85*(1), pp. 201-18.

_____ and Sul, Donggyu. "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel." *Journal of International Economics*, February 2001, *53*(1), pp. 29-52.

_____ and _____. "Asymptotic Power Advantages of Long Horizon Regressions." Unpublished manuscript, Ohio State University, April 2002.

_____; Ogaki, Masao and Sul, Donggyu. "Dynamic Seemingly Unrelated Cointegrating Regression." Unpublished manuscript, Ohio State University, 5 July 2000.

Marriott, F.H.C. and Pope, J.A. "Bias in the Estimation of Autocorrelations." *Biometrika*, December 1954, *41*(3/4), pp. 390-402.

Meese, Richard A. "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?" *Journal of Political Economy*, April 1986, *94*(2), pp. 345-73.

_____ and Rogoff, Kenneth. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, February 1983a, *14*(1/2), pp. 3-24.

_____ and _____. "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" in Jacob Frenkel, ed., *Exchange Rates and International Macroeconomics*. Chicago: NBER and University of Chicago Press, 1983b.

_____ and Rose, Andrew K. "Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation." *American Economic Review*, May 1990, *80*(2), pp.192-96.

Meredith, Guy and Chinn, Menzie D. ""Long-Horizon Uncovered Interest Rate Parity." Working Paper No. 6797, National Bureau of Economic Research, November 1998.

Mussa, Michael. "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating." *Scandinavian Journal of Economics*, 1976, *78*(2), pp. 229-48.

_____. "Empirical Regularities in the Behavior of Exchange Rates and Theories of the Foreign Exchange Market," in Karl Brunner and Allan H. Meltzer, eds., *Carnegie-Rochester Conference Series on Public Policy: Policies for Employment, Prices and Exchange Rates*, Vol. 11, 1979.

_____. "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications," in Karl Brunner and Allan H. Meltzer, eds., *Carnegie-Rochester, Conference Series on Public Policy: Real Business Cycles, Real Exchange Rates, and Actual Policies*, Vol. 25, Autumn 1986, pp. 117-214.

Neely, Christopher J. "Technical Analysis in the Foreign Exchange Market: A Layman's Guide." Federal Reserve Bank of St. Louis *Review*, September/October 1997, *79*(5), pp. 23-38.

_____; Weller, Paul; and Dittmar, Rob. "Is Technical

Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach." *Journal of Financial and Quantitative Analysis*, December 1997, *32*(4), pp. 405-26.

Pesaran, M. Hashem; Shin, Yongcheol and Smith, Ron P. "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels." *Journal of the American Statistical Association*, June 1999, *94*(446), pp. 621-34.

Phillips, Peter C.B. "Understanding Spurious Regressions in Econometrics." *Journal of Econometrics*, December 1986, *33*(3), pp. 311-40.

Rapach, David E. and Wohar, Mark E. "Testing the Monetary Model of Exchange Rate Determination: New Evidence from a Century of Data." Unpublished manuscript, November 2001a (forthcoming in *Journal of International Economics*, 2003, *59*(1), pp.1-2).

_____ and _____. "Testing the Monetary Model of Exchange Rate Determination: A Closer Look at Panels." Unpublished manuscript, September 2001b.

Rudebusch, Glenn D. "The Uncertain Unit Root in Real GNP." *American Economic Review*, March 1993, *83*(1), pp. 264-72.

Sarno, Lucio. "Toward a New Paradigm in Open Economy Modeling: Where Do We Stand?" Federal Reserve Bank of St. Louis *Review*, May/June 2001, *83*(3), pp. 21-36.

_____ and Taylor, Mark P. "The Microstructure of the Foreign Exchange Market: A Selective Survey of the Literature." *Princeton Studies in International Economics*, No. 89, April 2001.

_____ and _____. *The Economics of Exchange Rates*. Cambridge and New York: Cambridge University Press, 2002 (forthcoming).

Schinasi, Garry J. and Swamy, P.A.V.B. "The Out-of-Sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change." *Journal of International Money and Finance*, September 1989, *8*(3), pp. 375-90.

Stambaugh, Robert F. "Bias In Regressions With Lagged Stochastic Regressors." Working Paper No. 156, Center for Research in Security Prices, January 1986.

Stock, James H. and Watson, Mark W. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica*, July 1993, *61*(4), pp. 783-820.

Stockman, Alan C. "A Theory of Exchange Rate Determination." *Journal of Political Economy*, August 1980, *88*(4), pp. 673-98.

Taylor, Mark P. and Peel, David A. "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals." *Journal of International Money and Finance*, February 2000, *19*(1), pp. 33-53.

_____; _____ and Sarno, Lucio. "Nonlinear Mean-Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles." *International Economic Review* 2001, *42*(4), pp. 1015-42.

Wolff, Christian C.P. "Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models." *Journal of Business and Economic Statistics*, January 1987, *5*(1), pp. 87-97.

_____. "Exchange Rates, Innovations and Forecasting." *Journal of International Money and Finance*, March 1988, *7*(1), pp. 49-61.

Stock Market Returns, Volatility, and Future Output

Hui Guo

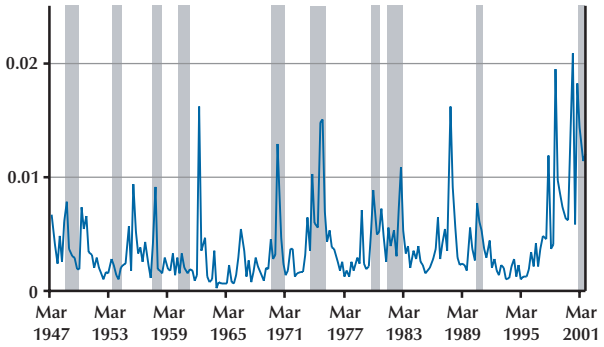
Stock market volatility is the systematic risk faced by investors who hold a market portfolio (e.g., a stock market index fund). Schwert (1989b) has undertaken an extensive study of stock market volatility, using historical data back to the 19th century. Some of his major findings are illustrated in Figure 1, which plots quarterly stock market volatility for the post-World War II period.[1] The figure shows that volatility moves countercyclically, exhibiting spikes during recessions. Also, stock market volatility tends to increase dramatically during financial crises (such as the 1987 stock market crash, the 1997 East Asia crisis, and the 1998 Russian bond default) and periods of uncertainty (such as the 1962 Cuban missile crisis). Moreover, volatility, once risen, shows some inertia in that it reverts only slowly to its previous, low level.

Although the causes of stock market volatility are not well understood, some authors suggest that elevated stock market volatility might reduce future economic activity.[2] Schwert (1989a) argues that stock market volatility, by reflecting uncertainty about future cash flows and discount rates, provides important information about future economic activity. Campbell et al. (2001), citing work by Lilien (1982), reason that stock market volatility is related to structural change in the economy. Structural change consumes resources, which depresses gross domestic product (GDP) growth. Another link between stock market volatility and output rests on a cost-of-capital channel. That is, an increase in stock market volatility raises the compensation that shareholders demand for bearing systematic risk. The higher expected return leads to the higher cost of equity capital in the corporate sector, which reduces investment and output. Consistent with these hypotheses about the link between stock market volatility and economic activity, Campbell et al. (2001) show that—after controlling for the lagged dependent variable—stock market volatility has significant predictive power for real GDP growth. Moreover, these authors also show that stock market volatility drives out returns in forecasting output. This finding deserves discussion.

Finance theory suggests that stock market returns rather than volatility have predictive power for investment and output because stock market returns are a forward-looking variable that incorporates expectations about future cash flows and discount rates. Several studies have confirmed the predictive power of stock market returns for investment and output, among them Fama (1981), Fischer and Merton (1984), and Barro (1990). On the other hand, the finding of Campbell et al. about the predictive power of stock market volatility for future economic activity is new, but the authors do not provide a theoretical explanation for the evidence. In this article, I try to reconcile the results of Campbell et al. with earlier empirical evidence on the predictive power of stock market returns and finance theory.

Figure 1

Realized Stock Market Variance



NOTE: Quarterly observations, 1947:Q2 through 2001:Q3; shaded areas indicate recessions dated by the National Bureau of Economic Research. See footnote 1 on adjustments to the October 1987 stock market crash.

Hui Guo is an economist at the Federal Reserve Bank of St. Louis. The author thanks Frank Schmid for helpful comments. William Bock provided research assistance.

[1] The stock market variance for 1987:Q4 amounts to several times more than the second-largest quarterly realization; this is due to a spike in volatility associated with the stock market crash on October 19, 1987. I follow Campbell et al. (2001) in replacing the variance of 1987:Q4 by the largest realized variance in the sample prior to 1987:Q4, and I make this adjustment throughout the paper. Appendix A describes the data.

[2] Shiller (1981) argues that stock prices are more volatile than what is justified by time variation in dividends. Similarly, Schwert (1989b) concludes that stock market volatility cannot be fully explained by changes in economic fundamentals.

I first use Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM) to illustrate the relation between stock market returns and volatility. I show that excess stock market returns, the difference between stock market returns and a risk-free rate, are positively correlated with one-period-lagged variance, but are negatively correlated with contemporaneous variance. These results have been well understood in the literature (e.g., Pindyck, 1988). Past variance relates positively to excess returns because it contains information about conditional variance or risk. The contemporaneous relation between excess returns and variance is negative because of a volatility feedback effect. That is, a positive innovation in variance today implies higher expected future variance and, therefore, higher expected future returns. For future expected returns to be higher, the innovation in variance must be accompanied by a drop in the stock market price index. Early authors (e.g., Pindyck, 1988; Turner, Startz, and Nelson, 1989; and Dueker, 1991) have found some support for this hypothesized risk-return relation. However, while Turner, Startz, and Nelson (1989) and Dueker (1991) impose no model restrictions on the coefficients of the variance terms, Pindyck (1988) finds that data reject these restrictions over some sample periods. In this paper, I find that (i) past stock market variance has significant forecasting ability for excess returns; (ii) the risk price is found to be positive and precisely estimated; and (iii) the model restrictions are not rejected by data.

After establishing the relation between excess returns and variance, it is straightforward to explain why stock market volatility drives out returns in forecasting output. According to the q theory of investment, an increase in stock market variance reduces investment—and hence output—contemporaneously because it raises the cost of capital. Lamont (2000), however, argues that investment expenditures react to changes in the cost of capital with lags. Therefore, stock market variance is negatively correlated with *future* investment and output. For the same reason, excess returns are expected to correlate positively with *future* output because excess returns correlate negatively with variance. It should be noted that excess returns—unlike variance—are hampered in their predictive power for future output because excess returns correlate positively with past variance, which in turn correlates negatively with future output. Because of these opposing effects, the predictive power of excess returns for future output is not as strong as the predictive power of variance. However, if the positive relation between excess returns and past variance is controlled for (i.e., adding past variance to the forecasting equation), excess returns might become significant and might even drive out variance because variance provides no additional information beyond excess returns in forecasting output.

I replicate the Campbell et al. (2001) result—that excess stock market returns are statistically insignificant in predicting GDP growth if stock market variance is also included in the forecasting equation—for the period 1963:Q1 to 1997:Q4. However, as postulated, I find that excess returns change from insignificant to marginally significant when I control for the lagged variance in the forecasting equation, meanwhile stock market variance changes from significant to marginally significant.[3] I also analyze two more sample periods. One sample covers the entire postwar period (1947:Q2 to 2000:Q4), while the other sample spans the longest available time period, ranging from 1885:Q4 to 2000:Q4. For these two sample periods, I find that excess returns actually drive out variance in forecasting output growth; moreover, only return terms are statistically significant if I also add past returns and past variance to the forecasting equation. Finally, the formal out-of-sample forecast test rejects the null hypothesis that excess returns provide no information about future GDP growth beyond what is contained in variance. These results should not be a surprise. As mentioned above, from the cost-of-capital point of view, volatility contains no additional information beyond excess returns; however, excess returns contain additional information (e.g., information about future cash flows) beyond variance in forecasting output.

In the article, I investigate the empirical link between stock market returns and volatility, analyze their relative forecasting power for output, and offer some concluding remarks.

EXCESS STOCK MARKET RETURNS AND VARIANCE

Merton (1973) shows that risk-averse investors demand extra compensation for bearing extra risk, everything else equal. Following Merton (1980), I assume that there is a linear positive relation between the stock market risk and return:

[3] If we extend the sample period to 2000:Q4, we find that the return term becomes significant while the variance term becomes insignificant.

(1) $$E_t e_{M,t+1} = \gamma E_t \sigma^2_{M,t+1},$$

where $E_t e_{M,t+1}$ is the conditional excess stock market return or the difference between the conditional return, $E_t r_{M,t+1}$, and a risk-free rate, $r_{f,t+1}$; $E_t \sigma^2_{M,t+1}$ is the conditional stock market variance; and $\gamma > 0$ is a measure of relative risk aversion. The degree of risk aversion can be interpreted as a price for risk.

Equation (1) holds exactly in the static Capital Asset Pricing Model (CAPM); however, in Merton's (1973) ICAPM, it holds only if investment opportunities are constant over time. In general, the expected return, $E_t e_{M,t+1}$, has an additional component reflecting the hedge demand for time-varying investment opportunities. Merton (1980) provides conditions under which the hedge component of excess returns is negligible, and equation (1) has been extensively utilized in the empirical literature.

Early work on the risk-return relation stated in equation (1) has come up with conflicting results. For example, while French, Schwert, and Stambaugh (1987) find a positive risk-return relation using an autoregressive conditional heteroskedasticity (ARCH) in mean model, Campbell (1987) documents a negative relation using an instrumental variable approach. Recent research by Scruggs (1998) and Guo (2002) suggests that the negative relation between stock market risk and return is an artifact of ignoring the hedge component of excess returns, which relates negatively to the risk component. After controlling for the hedge component in equation (1), these authors show that risk and return are indeed positively correlated. Moreover, Guo (2002) finds that high past stock market variance indeed forecasts high excess returns using quarterly data, although this relation is weakened by the hedge component. To focus on the issue of interest, I ignore the hedge component in this section and confine the analysis to the risk-return relation specified in equation (1), which is also the most widely used specification in the literature.

Existing empirical evidence suggests that stock market variance is serially correlated. To be parsimonious, I model the serial correlation as an AR(1) process, as in Pindyck (1988) and many others:

(2) $$\sigma^2_{M,t+1} = \alpha + \beta \sigma^2_{M,t} + \varepsilon_{t+1}.$$

Using the log-linearization method discussed in Campbell, Lo, and MacKinlay (1997), Guo and Whitelaw (2001) show that equations (1) and (2) imply

(3) $$e_{M,t+1} \approx \gamma\alpha + \gamma\beta\sigma^2_{M,t} - \frac{\gamma\rho\beta}{1-\rho\beta}\varepsilon_{t+1} + \eta_{d,t+1},$$

where ρ is a constant slightly less than 1 and $\eta_{d,t+1}$ is the shock to expected future dividend growth.[4] Equation (3) shows past variance can be used to forecast excess returns because past variance contains all the information about expected future variance and, hence, expected future excess return. It also shows that movements in excess returns are explained by past variance, shocks to variance, and shocks to dividend growth. In other words, the excess return at time $t+1$ is high because (i) expected return or variance at time t is high, (ii) there is a negative shock to variance at time $t+1$, and (iii) future dividend growth is expected to be higher than previously thought.

After rearranging terms, equation (3) can also be rewritten as

(4) $$e_{M,t+1} \approx \frac{\gamma \cdot \alpha}{1-\rho \cdot \beta} + \frac{\gamma\beta}{1-\rho\beta}\sigma^2_{M,t} - \frac{\gamma\rho\beta}{1-\rho\beta}\sigma^2_{M,t+1} + \eta_{d,t+1}.$$

Equation (4) shows that the excess return, $e_{M,t+1}$, is positively correlated with past variance, $\sigma^2_{M,t}$, and negatively correlated with contemporaneous variance, $\sigma^2_{M,t+1}$. As mentioned, the negative relation between excess returns and variance is caused by the serial correlation in variance or the volatility feedback effect.

Following Campbell et al. (2001), among many others, I measure stock market volatility by the sum of the squared deviations of the daily return on the market portfolio from its (daily) mean return. This volatility measure, as advocated by Merton (1980), is unbiased and can be arbitrarily accurate with sufficiently high-frequency data. I also experiment with some other volatility measures discussed in Yang and Zhang (2000), including the volatility estimator using high and low prices.[5] Interestingly, our measure outperforms the alternatives by a large margin in forecasting one-quarter-ahead excess returns. Moreover, although the alternative volatility measures differ somewhat from the one I use here, they never change the results in any qualitative ways. Many authors also use the volatility implied in stock market index options using the formula provided by Black and Scholes (1973). Unfortunately, historical

[4] Appendix B provides details of the deviation of equations (3) and (4).

[5] I use the daily high and low prices in Standard & Poor's 500 index.

Table 1

Stock Market Returns and Variance: OLS Regressions

	Constant	$\sigma^2_{M,t+1}$	$\sigma^2_{M,t}$	R^2
A. 1947:Q2–2000:Q4				
A1	0.001 (0.166)		4.367*** (2.633)	0.03
				0.03
A2	0.052*** (7.470)	–8.774*** (–4.144)		0.14
				0.14
A3	0.029*** (4.018)	–13.158*** (–5.287)	10.395*** (4.649)	0.29
				0.29
B. 1963:Q1–1997:Q4				
B1	–0.010 (–0.997)		6.487*** (3.554)	0.05
				0.04
B2	0.051*** (5.052)	–10.503*** (–3.564)		0.14
				0.13
B3	0.022** (2.366)	–18.427*** (–5.480)	16.112*** (6.640)	0.36
				0.35
C. 1885:Q4–2000:Q4				
C1	0.011** (2.358)		0.225 (0.581)	0.00
				–0.00
C2	0.026*** (3.842)	–2.011* (–1.830)		0.05
				0.05
C3	0.019*** (3.417)	–3.765*** (–4.054)	2.688*** (4.037)	0.10
				0.10

NOTE: t statistics are reported in parentheses and */**/*** denote significance at the 10/5/1 percent levels, respectively (two-tailed tests). For the column under R^2, the first row is R^2 and the second row is the adjusted R^2.

data on index options go back to the mid-1980s only. ARCH models are another popular method of estimating stock market volatility. Most appropriate for estimating volatility in high-frequency data, this concept of stock market volatility is less suitable for the purpose here.

I first estimate equations (3) and (4) using ordinary least squares (OLS). That is, I run OLS regressions of excess returns on past variance as in equation (3) and regress excess returns on both past and contemporaneous variance as in equation (4). It should be noted that OLS estimators are potentially biased for equation (4) because contemporaneous variance, $\sigma^2_{M,t+1}$, might be correlated with dividend shocks, $\eta_{d,t+1}$. Later, I report the generalized method of moments (GMM) estimation results, which are not vulnerable to the simultaneity problem. To control for serial correlation and heteroskedasticity in the residuals, I correct the standard errors following Newey and West (1987) when calculating t statistics.[6]

I analyze three different sample periods. Panel A in Table 1 represents the postwar sample, 1947:Q2 to 2000:Q4. As expected, although past variance exhibits a positive and statistically significant influence on excess returns in row A1, contemporaneous variance exhibits a negative and statistically significant correlation with excess returns in row A2. In the regression reported in row A3, both past and contemporaneous variance terms are statistically

[6] A procedure proposed by Newey and West (1994) is used to choose the appropriate lag length in constructing the covariance matrix.

Table 2

Stock Market Returns and Variance: GMM Regressions

	α	β	γ
A. 1947:Q2–2000:Q4			
	0.002*** (7.841)	0.499*** (8.829)	6.331*** (5.016)
R^2 of return equation: 0.28			
χ^2 (3) = 5.72 (p value 0.13)			
B. 1963:Q1–1997:Q4			
	0.001*** (6.100)	0.625*** (12.090)	5.868*** (4.971)
R^2 of return equation: 0.36			
χ^2 (3) = 5.64 (p value 0.13)			
C. 1885:Q4–2000:Q4			
	0.003*** (3.677)	0.502*** (3.847)	1.823** (2.148)
R^2 of return equation: 0.09			
χ^2 (3) = 6.15 (p value 0.11)			

NOTE: t statistics are reported in parentheses and */**/*** denote significance at the 10/5/1 percent levels, respectively (two-tailed tests). $\chi^2(3)$ is the statistic of over-identifying restrictions.

significant with expected signs. Interestingly, the R^2 statistic of 29 percent in row A3 is considerably greater than the R^2 statistics of rows A1 and A2. Also, the absolute values of coefficients and t statistics of the variance terms in row A3 are much larger than their counterparts in rows A1 and A2. This result can be attributed to the fact that—while variances are positively serially correlated—the effects of past and contemporaneous variance are of opposite signs in the excess return equation. Hence the evidence for a positive risk-return relation is stronger if the volatility feedback effect is controlled for.[7]

For the purpose of comparison, panel B reports the regression results for the sample period analyzed by Campbell et al. (2001), which runs from 1963:Q1 to 1997:Q4.[8] The results in panel B are similar to those in panel A, which is to be expected given the large overlap of the two sample periods. Lastly, panel C reports regression results for the most extensive available sample, 1885:Q4 to 2000:Q4. While the results in panel C are qualitatively similar to those in panels A and B, it should be noted that the variance terms have somewhat weaker explanatory power. Also, the absolute values of the coefficients of the variance terms are smaller than their counterparts in panels A and B.

As mentioned above, the OLS estimates of equation (4) reported in Table 1 might be biased because of a simultaneity problem. Also, it is more interesting to estimate the structural parameter or the price of risk, γ, rather than the coefficients of the variance terms. However, γ cannot be independently identified in equation (4). To address these issues, I use GMM to estimate equations (2) and (4) jointly[9]:

(5)

$$\sigma^2_{M,t+1} = \alpha + \beta\,\sigma^2_{M,t} + \varepsilon_{t+1}$$

$$e_{M,t+1} \approx \frac{\gamma\cdot\alpha}{1-\rho\cdot\beta} + \frac{\gamma\beta}{1-\rho\,\beta}\,\sigma^2_{M,t} - \frac{\gamma\rho\,\beta}{1-\rho\,\beta}\,\sigma^2_{M,t+1} + \eta_{d,t+1}.$$

The set of instrumental variables includes a constant and two lags in variance, which gives us six restrictions to identify three parameters, α, β, and γ. The system is thus overidentified with three degrees of freedom. The GMM regression results are reported

[7] Dueker (1991) makes a similar point by showing that the conditional variance is significantly positive (insignificantly negative) in the excess return equation if the contemporaneous variance is (not) controlled for.

[8] As in Campbell et al. (2001), I use the daily value-weighted market return data from the Center for Research in Security Prices (CRSP); the data span the period from 1962:Q3 to 1997:Q4. Two lags are used to forecast GDP; therefore, the analysis in Tables 1 through 3 uses data for the period 1963:Q1 to 1997:Q4.

[9] Following Campbell et al. (1997), I set ρ equal to 0.99 in quarterly data.

in Table 2; again, I analyze the aforementioned three sample periods. For all sample periods, all parameters are statistically significant. In particular, the risk price, γ, is positive and precisely identified.[10] Moreover, the overidentifying restrictions, which have a χ^2 distribution with three degrees of freedom, cannot be rejected at conventional significance levels.

EXCESS STOCK MARKET RETURNS, VARIANCE, AND FUTURE OUTPUT

Table 3 reports the OLS regression results of the one-quarter-ahead real GDP (GNP) growth rate, Δgdp_{t+1}, on lagged GDP growth, Δgdp_t; excess stock market return, $e_{M,t}$; variance, $\sigma^2_{M,t}$; and their lags.[11] Panel B shows the results for the Campbell et al. (2001) sample period 1963:Q1 to 1997:Q4. As shown in rows B1 and B2, both excess return, $e_{M,t}$, and variance, $\sigma^2_{M,t}$, have significant forecasting ability for the real GDP growth. The forecasting equation in row B3 includes lagged GDP growth, Δgdp_t, excess return, $e_{M,t}$, and variance, $\sigma^2_{M,t}$, which is the same specification used by Campbell et al. I confirm that variance drives out excess returns, while lagged GDP growth remains significant. As discussed above, these results may be explained by the positive correlation of excess return, $e_{M,t}$, with past variance, $\sigma^2_{M,t-1}$. To investigate this argument further, I add past variance to the forecasting equation, which leads to the regression results shown in row B4. Although insignificant, the coefficient of past variance is negative, as expected. Interestingly, now excess returns change from insignificant to marginally significant in row B3, while variance changes from significant to marginally significant in row B3. As a robustness check, I also add the past return, $e_{M,t-1}$, to the forecasting equation. The regression results of this approach are displayed in row B5, which shows that the past return is statistically significant, as is variance. Interestingly, if I extend the sample period to 2000:Q4, the return terms are always significant and the variance terms are always insignificant in the specifications of rows B4 and B5. To summarize, stock market variance drives out excess returns in forecasting output because of the positive relation between excess returns and past variance; if this relation is controlled for, excess returns show up significantly in the forecasting equation.

I repeat the analysis of the forecasting power of excess returns and variance for the two other samples mentioned. Panel A in Table 3 represents the full postwar sample. Again, as shown in rows A1 and A2, both excess return, $e_{M,t}$, and variance, $\sigma^2_{M,t}$, have significant forecasting ability for real GDP growth. However, in contrast with the finding in panel B, row A3 shows that excess returns actually drive out variance in the model specification chosen by Campbell et al. (2001). Consistent with the findings in panel B, row A4 shows that, when past variance is controlled for, both the coefficient and the t statistic of excess returns increase, while the corresponding values of variance decrease in absolute value terms. Row A4 also shows a statistically significant influence of past variance on future output. However, past variance loses its predictive power if I also include past returns in the forecasting equation: only return terms and lagged GDP growth have predictive power in row A5. Panel C reports the regression results using the sample 1885:Q4 to 2000:Q4. This long sample shows that return terms are always significant; also, variance terms are never significant except in row C1.

Table 3 provides in-sample evidence that excess stock market returns have significant predictive power for one-quarter-ahead GDP growth when I control for stock market variance and lagged GDP growth. The good in-sample fit of the return variable, however, does not guarantee its practical usefulness in output forecasts. To address this issue, I also perform out-of-sample forecast tests for the postwar sample and report the results in Table 4. In the benchmark model of Table 4 (panel A), GDP growth is a function of a constant, lagged GDP growth, variance, and past variance. To investigate whether excess returns provide additional information about future output beyond these variables, I augment the benchmark model by returns and past returns:

(6a)

$$\text{Benchmark: } dgdp_{t+1} = f(c, dgdp_t, \sigma^2_{M,t}, \sigma^2_{M,t-1})$$

$$\text{Augmented: } dgdp_{t+1} = f(c, dgdp_t, \sigma^2_{M,t}, \sigma^2_{M,t-1}, e_{M,t}, e_{M,t-1}).$$

Panel B proceeds in the other direction: including the return terms in the benchmark model and then testing whether the variance terms improve the forecast performance,

(6b)

$$\text{Benchmark: } dgdp_{t+1} = f(c, dgdp_t, e_{M,t}, e_{M,t-1})$$

$$\text{Augmented: } dgdp_{t+1} = f(c, dgdp_t, \sigma^2_{M,t}, \sigma^2_{M,t-1}, e_{M,t}, e_{M,t-1}).$$

[10] The price of risk is estimated to be slightly smaller here (panel A) than in Guo (2001) because the hedge component, which may have a nonzero mean, was included there. Interestingly, if a constant term is added to equation (4) as an additional parameter, the estimated risk price is close to that in Guo (2001).

[11] GNP data is used for the long sample only, which covers the period 1885:Q4 to 2000:Q4.

Table 3

Forecasting One-Quarter-Ahead GDP Growth

	Δgdp_t	$e_{M,t}$	$e_{M,t-1}$	$\sigma^2_{M,t}$	$\sigma^2_{M,t-1}$	R^2
A. 1947:Q2–2000:Q4						
A1	0.201** (2.396)			–0.952*** (–2.931)		0.16
						0.15
A2	0.330*** (6.198)	0.028*** (3.023)				0.16
						0.16
A3	0.320*** (5.979)	0.023*** (2.674)		–0.349 (–1.573)		0.17
						0.16
A4	0.297*** (4.967)	0.030*** (3.503)		–0.102 (–0.398)	–0.430** (–2.092)	0.19
						0.17
A5	0.257*** (4.434)	0.024*** (3.039)	0.035*** (3.290)	–0.222 (–0.934)	–0.077 (–0.425)	0.24
						0.23
B. 1963:Q1–1997:Q4						
B1	0.322*** (5.892)			–0.547** (–2.073)		0.15
						0.14
B2	0.293*** (3.399)	0.024** (2.506)				0.13
						0.12
B3	0.218** (2.343)	0.014 (1.355)		–0.783** (–2.383)		0.18
						0.16
B4	0.200* (1.879)	0.020* (1.777)		–0.551* (–1.661)	–0.364 (–0.905)	0.18
						0.16
B5	0.179* (1.891)	0.014 (1.476)	0.027** (2.300)	–0.615** (–2.173)	–0.017 (–0.050)	0.23
						0.21
C. 1885:Q4–2000:Q4						
C1	0.368*** (4.281)			–0.254*** (–2.824)		0.17
						0.17
C2	0.330*** (3.525)	0.070*** (5.669)				0.25
						0.24
C3	0.313*** (2.910)	0.068*** (5.172)		–0.151 (–1.358)		0.25
						0.25
C4	0.317*** (2.937)	0.066*** (5.156)		–0.219 (–1.530)	0.100 (0.663)	0.25
						0.24
C5	0.344*** (3.617)	0.064*** (5.292)	–0.017 (–0.955)	–0.209 (–1.408)	0.074 (0.472)	0.26
						0.25

NOTE: t statistics are reported in parentheses and */**/*** denote significance at 10/5/1 percent levels, respectively (two-tailed tests). For the column under R^2, the first row is R^2 and the second row is the adjusted R^2.

Table 4

One-Quarter-Ahead Forecasts of GDP Growth: Nested Comparisons

	ENC-NEW		MSE-F	
MSE_A / MSE_B	Statistic	Asy. CV	Statistic	Asy. CV
A. Variance in benchmark model				
0.94	24.06	2.88	9.17	1.71
B. Return in benchmark model				
1.02	2.06	2.88	–2.62	1.71

NOTE: MSE_A (MSE_B) is the mean-squared-error of the augmented (benchmark) model. See Appendix C for details about the ENC-NEW and MSE-F tests.

Three statistics are calculated to compare the performance of the augmented model with the performance of the respective benchmark model. The first statistic is the mean-squared-error ratio. The second one is the encompassing test (ENC-NEW) developed by Clark and McCracken (1999). The ENC-NEW statistic tests the null hypothesis that the benchmark model encompasses all the relevant information for one-quarter-ahead GDP growth; the alternative hypothesis of panel A (panel B) states that the return (variance) variables provide additional information. The third statistic is the equal forecast accuracy test (MSE-F) developed by McCracken (1999). The null hypothesis states that the benchmark model has a mean-squared forecasting error less than or equal to the error of the augmented model; the alternative is that the augmented model has a smaller mean-squared error. Clark and McCracken (1999) show that these two tests have the best overall power and size properties. I provide some details about these two tests in Appendix C. The columns titled "Asy. CV" report the 95 percent asymptotic critical values provided by Clark and McCracken (1999) and McCracken (1999). I use the period 1947:Q3 to 1965:Q1 for the initial in-sample regression and use the regression results to forecast GDP growth in 1965:Q2. The in-sample regression is updated recursively. That is, I then use the period 1947:Q3 to 1965:Q2 to make a forecast for the GDP growth in 1965:Q3 and so on. Panel A shows that incorporating the return variables reduces the forecasting error. The mean-squared-error of the augmented model amounts to only 94 percent of the mean-squared-error of the benchmark model. Also, the ENC-NEW and MSE-F tests reject the null hypotheses overwhelmingly. Hence, the out-of-sample forecast tests in panel A provide strong evidence that excess returns contain information about future output beyond variance. On the other hand, panel B shows that the augmented model exhibits a higher mean-squared-error than the benchmark model. Moreover, in panel B the null hypotheses are not rejected at the conventional significance levels for either the ENC-NEW or the MSE-F tests. Therefore, I cannot reject the null hypothesis that returns subsume the information content of variance in forecasting real output growth.

CONCLUSION

In this article, I show that there is a close link between stock market returns and volatility. That is, because volatility is serially correlated, returns relate positively to past volatility, but relate negatively to contemporaneous volatility. Therefore, stock market volatility forecasts output because volatility affects the cost of capital through its link with expected stock market return. From the cost-of-capital point of view, volatility contains no additional output-forecasting information beyond the information that returns provide, although the positive relation between returns and past volatility weakens the predictive power of returns in certain specifications. On the other hand, stock market returns do contain information about future economic activity beyond volatility (e.g., information about future cash flows). Therefore, if the cost of capital is the main channel through which volatility affects future output, it should follow that stock market returns have a more important role in forecasting economic activity than volatility does. I show that this hypothesis is supported by the in-sample and out-of-sample regression results using postwar data.

REFERENCES

Barro, Robert J. "The Stock Market and Investment." *Review of Financial Studies*, 1990, *3*(1), pp. 115-31.

Black, Fischer, and Scholes, Myron S. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, May/June 1973, *81*(3), pp. 637-54.

Campbell, John Y. "Stock Returns and the Term Structure." *Journal of Financial Economics*, June 1987, *18*(2), pp. 373-99.

_____ ; Lettau, Martin; Malkiel, Burton G. and Xu, Yexaio. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance*, February 2001, *56*(1), pp. 1-43.

_____ ; Lo, Andrew and MacKinlay, A. Craig. *The Econometrics of Financial Markets*. Princeton: Princeton University Press, 1997.

Clark, Todd E. and McCracken, Michael W. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." Working Paper No. 99-11, Federal Reserve Bank of Kansas City, October 1999.

Cochrane, John H. "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations." *Journal of Finance*, March 1991, *46*(1), pp. 209-37.

Dueker, Michael J. "The News Effect of Time-Varying Kurtosis on Stock Returns." Working paper, University of Washington, 1991.

Fama, Eugene F. "Stock Return, Real Activity, Inflation, and Money." *American Economic Review*, September 1981, *71*(4), pp. 545-65.

Fisher, Stanley and Merton, Robert C. "Macroeconomics and Finance: The Role of the Stock Market." Carnegie-Rochester Conference Series on Public Policy, Autumn 1984, *21*(0), pp. 57-108.

French, Kenneth R.; Schwert, G. William and Stambaugh, Robert F. "Expected Stock Returns and Volatility." *Journal of Financial Economics*, September 1987, *19*(1), pp. 3-30.

Gordon, Robert J. *The American Business Cycle*. Chicago: The University of Chicago, 1986.

Guo, Hui. "Understanding the Risk-Return Tradeoff in the Stock Market." Working Paper No. 2002-001A, Federal Reserve Bank of St. Louis, 2002.

_____ and Whitelaw, Robert. "Risk and Return: Some New Evidence." Working Paper 2001-A, Federal Reserve Bank of St. Louis, 2001.

Lamont, Owen A. "Investment Plans and Stock Returns." *Journal of Finance*, December 2000, *55*(6), pp. 2719-45.

Lilien, David M. "Sectoral Shifts and Cyclical Unemployment." *Journal of Political Economy*, August 1982, *90*(4), pp. 777-93.

McCracken, Michael W. "Asymptotics for Out-of-Sample Tests of Causality." Working paper, Louisiana State University, 1999.

Merton, Robert C. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, September 1973, *41*(5), pp. 867-87.

_____. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics*, December 1980, *8*(4), pp. 323-61.

Newey, Whitney K. "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, May 1987, *55*(3), pp. 703-08.

_____ and West, Kenneth D. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, October 1994, *61*(4), pp. 631-54.

Pindyck, Robert S. "Risk Aversion and Determinants of Stock Market Behavior." *Review of Economics and Statistics*, May 1988, *70*(2), pp. 183-90.

Schwert, G. William. "Business Cycles, Financial Crises, and Stock Volatility." Carnegie-Rochester Conference Series on Public Policy, Autumn 1989a, *31*(0), pp. 83-126.

_____. "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, December 1989b, *44*(5), pp. 1115-53.

_____. "Indexes of U.S. Stock Prices from 1802 to 1987." *Journal of Business*, July 1990, *63*(3), pp. 399-426.

Scruggs, John T. "Resolving the Puzzling Intertemporal Relation Between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach." *Journal of Finance*, April 1998, *53*(2), pp. 575-603.

Shiller, Robert J. "Do Stock Prices Move Too Much To Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, June 1981, *71*(3), pp. 421-36.

Turner, Christopher; Startz, Richard and Nelson, Charles R. "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market." *Journal of Financial Economics*, November 1989, *25*(1), pp. 3-22.

Yang, Dennis, and Zhang, Qiang. "Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices." *Journal of Business*, July 2000, *73*(3), pp. 477-91.

Appendix A

DATA DESCRIPTION

The daily excess stock market return is the difference between the daily stock market return and the daily risk-free rate. I use the daily market return constructed by Schwert (1990) before July 2, 1962, and use the daily value-weighted stock market return obtained from the Center for Research in Security Prices (CRSP) through the end of the year 2000; thereafter I use the daily S&P 500 index. The risk-free rate is the commercial paper rate from Appendix B of Gordon (1986) before 1926 and is obtained from CRSP thereafter. The risk-free rate is available only at a monthly frequency. I calculate a daily risk-free rate from monthly observations by dividing by the number of trading days in the respective month. Thus I assume that the daily risk-free rate is constant within each month. I aggregate the daily excess return to quarterly observation, $e_{M,t}$. The stock market variance, $\sigma_{M,t}$, is calculated as the sum of squared deviations from the mean of daily excess stock market returns:

$$\sigma_{M,t} = \sum_{j=1}^{\tau} (e_{M,t\tau} - \bar{e}_{M,t})^2, \tag{A1}$$

where $e_{M,t\tau}$ is the excess return of day τ in quarter t and $\bar{e}_{M,t}$ is the average daily excess return of quarter t. The quarterly real GDP data are obtained from the Bureau of Economic Analysis (BEA). The data of real GNP before 1947:Q1 are from Appendix B of Gordon (1986) and from the BEA thereafter. The growth rates of GDP (GNP) are calculated as differences in logarithmic values (log growth rates). Source: CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago, 2002. Used with permission. All rights reserved. < www.crsp.uchicago.edu > .

Appendix B

A LOG-LINEAR ASSET PRICING MODEL

Using a log-linearization method, Campbell et al. (1997) show that excess stock market return, $e_{M,t+1}$, can be approximately expressed as

(A1)
$$e_{M,t+1} \approx E_t e_{M,t+1} + (E_{t+1} - E_t)$$
$$[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - \sum_{j=1}^{\infty} \rho^j \Delta e_{M,t+1+j} - \sum_{j=1}^{\infty} \rho^j r_{f,t+1+j}],$$

where Δd_{t+1+j} is the growth rate of the dividend, $r_{f,t+1+j}$ is the real risk-free rate, and ρ is a constant. Substituting equation (1) into (A1), I obtain

(A2)
$$e_{M,t+1} \approx \gamma E_t \sigma^2_{M,t+1} - (E_{t+1} - E_t)[\sum_{j=1}^{\infty} \rho^j \sigma^2_{M,t+1+j}] + \eta_{d,t+1} + \eta_{f,t+1},$$

where

$$\eta_{d,t+1} = (E_{t+1} - E_t)[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j}]$$

and

$$\eta_{f,t+1} = -(E_{t+1} - E_t)[\sum_{j=1}^{\infty} \rho^j r_{f,t+1+j}].$$

Substituting equation (2) into equation (A2), I obtain

(A3)
$$e_{M,t+1} \approx \gamma\alpha + \gamma\beta\sigma^2_{M,t} - \frac{\gamma\rho\beta}{1-\rho\beta}\varepsilon_{t+1} + \eta_{d,t+1} + \eta_{f,t+1}.$$

I can rewrite equation (2) as

(2)′
$$\varepsilon_{t+1} = \sigma^2_{M,t+1} - \alpha - \beta\sigma^2_{M,t}.$$

Substituting equation (2)′ into equation (A3), I obtain

(A4)
$$e_{M,t+1} \approx \frac{\gamma\cdot\alpha}{1-\rho\cdot\beta} + \frac{\gamma\beta}{1-\rho\beta}\sigma^2_{M,t} - \frac{\gamma\rho\beta}{1-\rho\beta}\sigma^2_{M,t+1} + \eta_{d,t+1} + \eta_{f,t+1}.$$

Campbell et al. (1997) show that shocks to the real risk-free rate, $\eta_{f,t+1}$, explain little of the variation in excess stock market returns. By setting $\eta_{f,t+1}$ equal to zero, (A3) and (A4) become equations (3) and (4), respectively.

Appendix C

ENCOMPASSING (ENC-NEW) TEST AND EQUAL FORECAST ACCURACY (MSE-F) TEST

Suppose that there are $T+1$ observations. The first R observations are used in the initial in-sample regression. The forecasting error for period $R+1$ is $\hat{u}_{B,R+1}$ for the benchmark model and $\hat{u}_{A,R+1}$ for the augmented model. Then estimates of the forecasting equations are updated using the first $R+1$ observations and used to forecast the next period, $R+2$. The forecasting error for period $R+2$ is $\hat{u}_{i,R+2}$, $i = A, B$. Forecasts are recursively updated to generate a time series of one-period-ahead forecasting errors $\{\hat{u}_{i,\tau}\}$, $i = A, B$ and $\tau = R+1 \ldots T+1$, a total of $P = T+1-R$ observations.

The statistic of the ENC-NEW test is

$$ENC-NEW = P\frac{P^{-1}\sum_{t=R}^{T}[\hat{u}^2_{B,t+1} - \hat{u}_{B,t+1}\hat{u}_{A,t+1}]}{P^{-1}\sum_{t=R}^{T}\hat{u}^2_{A,t+1}}.$$

Clark and McCracken (1999) derive the asymptotic distribution of the ENC-NEW statistic under the null hypothesis that the augmented model encompasses the information of the benchmark model.

The MSE-F test statistic is

$$MSE-F = \frac{\sum_{t=R}^{T}[\hat{u}^2_{B,t+1} - \hat{u}^2_{A,t+1}]}{P^{-1}\sum_{t=R}^{T}\hat{u}^2_{A,t+1}}.$$

McCracken (1999) derives the asymptotic distribution of the MSE-F statistic under the null hypothesis that the augmented model has a smaller mean-squared forecasting error than the benchmark model does.

All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* are available to our readers on our Web site: **<research.stlouisfed.org/publications/review>**. Also, you may request data and programs on either disk or hard copy: **Research Department**, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442. Please include the author, title, issue date, and page numbers with your request.

These data and programs are also available through **Inter-university Consortium for Political and Social Research (ICPSR)**. Member institutions may request data through the CDNet Order facility. Nonmembers may write to ICPSR, Institute for Social Research, P.O. Box 1248, Ann Arbor, MI 48106-1248; call 734-998-9900; or e-mail **<netmail@icpsr.umich.edu>**.

General data can be obtained through **FRED (Federal Reserve Economic Data)**, a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our Web site: **<research.stlouisfed.org/fred>**.

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442