Federal Reserve Bank of New York

December 2007
Volume 13 Number 3

Economic Policy Review

The *Economic Policy Review* is published by the Research and Statistics Group of the Federal Reserve Bank of New York. The views expressed are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

www.newyorkfed.org/research

Federal Reserve Bank of New York
Economic Policy Review

December 2007
Volume 13 Number 3

Contents

Articles:

Hedge funds, with assets under management approaching an estimated $1.5 trillion in 2006, have become important players in the U.S. and global capital markets. These largely unregulated funds differ from other market participants in their use of a variety of complex trading strategies and instruments, in their liberal use of leverage, in their opacity to outsiders, and in their convex compensation structure. These differences can exacerbate market failures associated with agency problems, externalities, and moral hazard. Counterparty credit risk management (CCRM) practices, used by financial institutions to assess credit risk and limit counterparty exposure, are the first line of defense against market disruptions with potential systemic consequences. This article examines how the unique nature of hedge funds may generate market failures that make CCRM for exposures to the funds intrinsically more difficult to manage, both for regulated institutions and for policymakers concerned with systemic risk. The authors acknowledge that various market failures, such as the one linked to the 1998 collapse of hedge fund Long-Term Capital Management, may make CCRM imperfect. However, CCRM has improved significantly since then, and it remains the appropriate starting point for limiting the potential for hedge funds to generate systemic disruptions.

19 A Comparison of Measures of Core Inflation

Robert Rich and Charles Steindel

The ability of central banks to differentiate between permanent and transitory price movements is critical for the conduct of monetary policy. The importance of gauging the persistence of price changes in a timely manner has led to the development of measures of underlying, or "core," inflation that are designed to remove transitory price changes from aggregate inflation data. Given the usefulness of this information to policymakers, there is a surprising lack of consensus on a preferred measure of U.S. core inflation. This article examines several proposed measures of core inflation—the popular ex food and energy series, an ex energy series, a weighted median series, and an exponentially smoothed series—to identify a "best" measure. The authors evaluate the measures' performance according to criteria such as ease of design and accuracy in tracking trend inflation, as well as explanatory content for within-sample and out-of-sample movements in aggregate CPI and PCE inflation. The study reveals that the candidate series perform very differently across aggregate inflation measures, criteria, and sample periods. The authors therefore find no compelling evidence to focus on one particular measure of core inflation, including the series that excludes food and energy prices. They attribute their results to the design of the individual measures and the measures' inability to account for variability in the nature and sources of transitory price movements.

39 The Role of Retail Banking in the U.S. Banking Industry: Risk, Return, and Industry Structure

Timothy Clark, Astrid Dick, Beverly Hirtle, Kevin J. Stiroh, and Robard Williams

The U.S. banking industry is experiencing a renewed interest in retail banking, broadly defined as the range of products and services provided to consumers and small businesses. This article documents the "return to retail" in the U.S. banking industry and offers some insight into why the shift has occurred. At the bank level, the principal attraction of retail banking seems to be the belief that its revenues are stable and thus can offset volatility in nonretail businesses. At the industry level, the authors show that interest in retail activities fluctuates in rather predictable ways with the performance of nonretail banking and financial market activities. They document the features that the recent "return to retail" has in common with past cycles, but also identify factors suggesting that this episode may be more persistent. The most important of these factors is the role of large banks: this retail banking cycle is being driven almost entirely by the very largest U.S. banking firms. The key role of very large banks gives extra weight to this retail banking episode.

John Kambhu, Til Schuermann, and Kevin J. Stiroh

Hedge Funds, Financial Intermediation, and Systemic Risk

- An important channel through which largely unregulated hedge funds interact with regulated institutions is prime brokerage relationships.
- Central to these relationships is the extension of credit to hedge funds, which exposes banks to counterparty credit risk.
- Counterparty credit risk management (CCRM) practices, used to assess credit risk and limit counterparty exposure, are banks' first line of defense against market disruptions with potential systemic consequences.
- Hedge funds' unrestricted trading strategies, liberal use of leverage, opacity to outsiders, and convex compensation structure make CCRM more difficult, as they exacerbate potential market failures.
- While past market failures suggest that CCRM is not perfect, it remains the best initial safeguard against systemic risk; thus, the current emphasis on CCRM as the primary check on hedge fund risk-taking is appropriate.

1. Introduction

Financial economists and policymakers have historically focused on banks as prospective channels of systemic distress through, for instance, bank runs and the concomitant reduction in the supply of credit. This "special" attribute of banks has been behind the classic policy rationale for regulating them. The ongoing move toward financial markets, arm's-length transactions, and active trading, however, has shifted focus to the potential impact of a hedge-fund-led disruption on financial institutions, markets, and the broader economy.[1]

Financial intermediaries, of course, have many ways to reduce their exposure and mitigate the impact of financial market shocks. The first line of defense is the intermediary's counterparty credit risk management (CCRM) system. Banks establish limits; implement risk reporting infrastructures; and define haircut, margining, and collateral policies—all designed to assess credit risk and limit their counterparty exposure. Effective CCRM is obviously needed for any counterparty, but hedge funds differ in important ways, such as in their use of

[1] See, for example, McCarthy (2006), President's Working Group on Financial Markets (2007), and the papers in the Banque de France (2007) special issue devoted to hedge funds. In addition to concerns about financial system implications, there are concerns about investor protection and market integrity issues, which we do not discuss.

John Kambhu is a vice president and Til Schuermann an assistant vice president at the Federal Reserve Bank of New York; Kevin J. Stiroh was a vice president at the Bank at the time this article was written.
<john.kambhu@ny.frb.org>
<til.schuermann@ny.frb.org>

The authors thank Tobias Adrian, Art Angulo, Kevin Coffey, Mark Flannery, Beverly Hirtle, Steve Manzari, Trish Mosser, Mike Nelson, Brian Peters, Asani Sarkar, Phil Strahan, Zhenyu Wang, two anonymous referees, and seminar participants at the Federal Reserve Bank of New York for helpful suggestions, and Sarita Subramanian for excellent research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

complex trading strategies and instruments, leverage, opacity, and convex compensation structures, all of which increase the challenges to effective CCRM.

This article examines how the nature and characteristics of hedge funds may generate "market failures" that make CCRM for exposures to hedge funds intrinsically more difficult to manage, both for the individual firm and for policymakers concerned with systemic risk. We put forward no specific new policy proposals, however, because we believe CCRM remains the appropriate starting point for limiting the potential for hedge funds to generate systemic disruptions.[2] By laying out the issues and highlighting the specific linkages from hedge funds to systemic risk, we hope to highlight areas for further research on when and how markets may fail to yield a desirable outcome.

2. Hedge Funds 101

We begin by describing the difference between a hedge fund and other asset management vehicles such as mutual or pension funds, then discuss the traditional role of counterparty credit risk management, and present some stylized facts about the hedge fund industry.

2.1 What Is a Hedge Fund?

Hedge funds, in short, are largely unregulated, private pools of capital. Hedge fund managers can invest in a broad array of assets and pursue many investment strategies, such as global macro, market neutral equity, convertible arbitrage, or event-driven.[3] While strategies and individual hedge funds are quite heterogeneous, it is useful to focus on four broad characteristics that distinguish hedge funds from other types of money management funds.

First, hedge funds are not restricted by the type of trading strategies and financial instruments they may use. In particular, hedge funds can and do make use of short-selling, derivatives, and options, all of which are complex and potentially nonlinear in payoffs. Second, hedge funds make liberal use of leverage, be it directly through the use of debt or indirectly through leverage embedded in derivatives. This freedom is possible because hedge funds in the United States largely fall outside of the regulatory umbrella by virtue of being open only to accredited investors and large institutions.[4] Of course, hedge fund investors and counterparties impose some discipline on the amount of leverage actually employed. This discipline, however, may be limited by the third key characteristic—opacity to outsiders—which again is in large part due to the funds' unregulated nature. Finally, hedge fund managers are typically compensated based on both scale and absolute performance through a dual fee structure, for example, the "2-and-20" set-up whereby managers retain 2 percent of the net asset value of the fund and 20 percent of returns in excess of some benchmark.

Hedge funds, in short, are largely unregulated, private pools of capital.

The first two strategies—the use of complex, nonlinear financial instruments and the use of leverage—make hedge funds somewhat unusual in the asset management world, but not unique among financial intermediaries.[5] Mutual funds, for example, have a limited ability to short-sell.[6] In contrast, both of these strategies are available to commercial and investment banks, and many have proprietary trading units that emulate hedge fund investment strategies. This is important, as those firms often act as counterparties to hedge funds and likely have experience with a range of financial instruments and strategies, and should therefore be able to adjust their risk management practices accordingly. It is precisely this flexibility, however, that allows hedge funds to play their critical role in terms of price discovery, arbitrage, and increased market efficiency.

Opacity also is not unique to hedge funds as financial institutions generally, and banks especially are thought to be more opaque than firms in other industries.[7] Financial firms, however, often have traded instruments such as equity shares or bonds outstanding, so they are subjected to further scrutiny by

[2]Supervisors, of course, have other tools such as direct regulation or disclosure requirements that may mitigate the potential for systemic disruptions. We discuss these in Section 5.

[3]See the useful overviews by Fung and Hsieh (1999), McCarthy (2006), Hildebrand (2007), and Stulz (2007).

[4]The regulatory scope of the Securities and Exchange Commission is restricted to protecting small, retail investors. The term "accredited investor" helps define what is in and out of scope. See <http://www.sec.gov/answers/accred.htm>.

[5]There are other investment vehicles that make use of some of these strategies, such as private equity and vulture funds.

[6]See Almazan et al. (2004) for a description of the regulations that govern how mutual funds may behave in terms of issuing shares, distributing dividends, and reporting, as well as investment restrictions such as the use of leverage and short-selling.

[7]Morgan (2002) shows that rating agencies disagree more about financial firms than about nonfinancials, a sign of opacity, although Flannery, Kwan, and Nimalendran (2004) present an opposing view.

market participants. In contrast, hedge funds, due to their unregulated and private nature, are not subject to such wide scrutiny unless they choose to issue public securities. Moreover, the success of a hedge fund often depends on proprietary trading strategies that, if made public, can be used by others to trade against them. Investors know this and are thus willing to tolerate a degree of opacity not seen in the mutual fund industry in the hope of securing particularly rich returns.[8]

Finally, the compensation structure of hedge funds differs markedly from that of other institutional investors such as mutual funds. In particular, hedge fund traders and managers tend to be compensated more on absolute return and scale, while their brethren in institutional investing typically have their compensation tied to performance relative to some benchmark such as the S&P 500.[9] Hedge fund managers also have added optionality in the form of hurdle rates (no incentive fee if returns are below the hurdle rate) and high-water marks (incentive fees only on new profits, that is, after past losses are made up), making payoffs potentially very convex and therefore asymmetric: gains and losses are treated differently (Ackermann, McEnally, and Ravenscraft 1999).[10] This convex payoff structure provides strong incentives for hedge fund managers to take on risk and leverage. Although incentive fees also play an important role in the mutual fund industry, they are required by law to be symmetric in the United States (Elton, Gruber, and Blake 2003).

Investors prefer hedge fund managers with a significant fraction of their personal wealth invested in the fund . . . but information on manager stakes is difficult to come by and is generally known only to the largest and most sophisticated investors.

The combination of opacity and highly convex compensation structures has the potential to create excess risk-taking, which may make investors reluctant to commit capital. As a result, investors prefer hedge fund managers with a significant fraction of their personal wealth invested in the fund, that is, "skin in the game," but information on manager stakes is difficult to come by and is generally known only to the largest and most sophisticated investors. Consequently, a hedge fund's ability to raise and retain capital may be particularly acute in periods of stress, when investors may try to withdraw capital from hedge funds.

To be sure, these characteristics are relevant for many types of financial firms, such as the proprietary trading desks at regulated financial institutions. Hedge funds are arguably different in degree but not in kind; thus, any preventive measures or policy discussions should not be limited to hedge funds alone. We now turn to traditional tools used to reduce exposure.

2.2 Counterparty Credit Risk Management

Hedge funds interact with regulated financial institutions and intermediaries in many ways, including prime brokerage relationships, where regulated intermediaries provide services such as trading and execution, clearance and custody, securities lending, technology, and financing through margin loans and repurchase agreements.[11] An important part of these relationships is the extension of credit to the hedge fund, so the financial institution is exposed to counterparty credit risk. As a result, traditional CCRM systems are the first line of defense between unregulated hedge funds and regulated financial institutions.

An integral part of CCRM is margining and collateral practices, which are designed to reduce counterparty credit risk in leveraged trading by providing a buffer against increased exposure to the dealer providing the financing or derivatives contract. In general, a financial institution may be willing to extend credit to the hedge fund against the posting of specific collateral that is valued at no less than the amount of the exposure. This reduction in settlement risk in leveraged trading increases confidence and thereby promotes active financing of leveraged trading.

To be precise, variation margin is the amount of collateral or cash provided to a dealer to cover past changes in the value of theprices move and evolve, for example, the mark-to-market of the position may deteriorate and trigger a margin call. Initial margin, in contrast, is an additional amount of collateral designed to cover potential future changes in the value of the contract (potential future exposure). Variation margin and

[8]This increased managerial discretion does result in higher risk-adjusted returns, as found by Agarwal, Daniel, and Naik (2007).

[9]Relative returns do matter to the extent that capital flows toward better performing hedge funds. That is, high-water marks compensate absolute returns only to the extent that capital under management is fixed. As we discuss later, hedge fund capital does have some mobility.

[10]Agarwal, Daniel, and Naik (2007) report that more than 60 percent of funds have hurdle rates and more than 80 percent have high-water-mark provisions. Note that these authors find that funds with high-water marks earn higher returns—about 20 percent higher—as they should, but hurdle rates have no effect on fund performance.

[11]Revenues for prime brokerage services alone accounted for about $8 billion in 2005, with total hedge-fund-related business revenue estimated at $26 billion, about 2 percent of the funds' total assets under management (Richard Beales and Joana Chung, "Banks Take to a Supporting Role as Hedge Funds Flourish," *Financial Times*, August 9, 2006, p. 7).

Box 1

Determination of Initial Margins

Consider the initial margin standards of exchanges such as the Chicago Mercantile Exchange or the Chicago Board of Trade. Specifically, the March 12, 2007, initial margin requirement for a JPY/USD futures contract is $1,350.[a] These contracts are sold in units of ¥12,500,000 each. On March 12, 2007, the exchange rate was $1 = ¥116. The annual volatility for this exchange rate is measured to be about 8.4 percent. Then, one-day 99 percent VaR turns out to be very close to the initial margin requirement:

$$VaR = 2.33 \times \left(\frac{8.4\%}{\sqrt{252}}\right) \times \left(\frac{12{,}500{,}000}{116}\right) = \$1{,}300,$$

where the annual volatility is converted to a daily volatility using the standard approach of normalizing by (the square root of) the number of trading days in a year. This exercise makes clear that risk-based initial margin requirements depend on market conditions, here the exchange rate and its volatility. If market conditions worsen, the initial margin would be too low and additional margin would be called for.

The level of initial margin can be more sensitive to volatility and market conditions in over-the-counter (OTC) markets, where trading terms are more flexible than on futures exchanges. Nevertheless, even in OTC markets, margin locks are being adopted to provide more predictability of initial margin requirements. In a margin-lock, a dealer will commit to freeze the initial margin terms (for example, the volatility term in the formula above) for a specified time period, say, three months. This practice reduces the liquidity risk for the trading counterparty—though at the cost to the dealer of having fixed initial margin terms at a time when the volatility of exposures may be rising. Note, however, that margin locks do not preclude the dealer from collecting variation or maintenance margins.

[a]This example is adapted from Jorion (2007). Margin information for the Chicago Mercantile Exchange can be found at <http://www.cme.com/html.wrap/wrappedpages/clearing/pbrates/PBISHomePage.htm>.

initial margin together ensure that collateral held by the dealer is sufficient to cover the current replacement value of the contract if the counterparty defaults, and also the potential change in value of the contract between the time of the default and the time at which the trading position can be liquidated. Initial margins vary by financial instrument and are usually set to cover changes in the contract's value up to a certain probability, typically 95 percent to 99 percent, over a particular horizon, typically one day to two weeks (see Box 1).

Other forms of traditional CCRM include the development of a broad set of risk metrics including internal ratings; ongoing monitoring and evaluation of exposures such as stress testing on a consolidated basis over a range of suitably stressful scenarios; due diligence to understand the strategies and history of the counterparty; limits on specific trades, exposures, or concentrations; and well-defined processing arrangements and settlement protocols. All of these help control exposure and reduce the risk of the financial institution dealing with a hedge fund counterparty.

Our discussion so far has focused on the CCRM imposed by the financial intermediaries that interact directly with a counterparty such as a hedge fund. Of course, investors also play a critical role in disciplining hedge fund behavior and reducing excess risk-taking. It is clearly in the interest of individual investors to understand and evaluate the objectives, strategies, fee structures, and history of the particular funds in which they invest, but the same factors that make CCRM difficult also increase the challenges associated with investor market discipline.

A recent industry study, however, reported improvements in risk management practices in the global hedge fund industry and noted in particular that 87 percent of dealers surveyed were actively negotiating credit terms specifically to increase transparency and disclosure.[12] Moreover, the invested capital of the hedge fund managers and the managers' desire to maintain the franchise value of the fund also provide clear incentives to improve risk management.

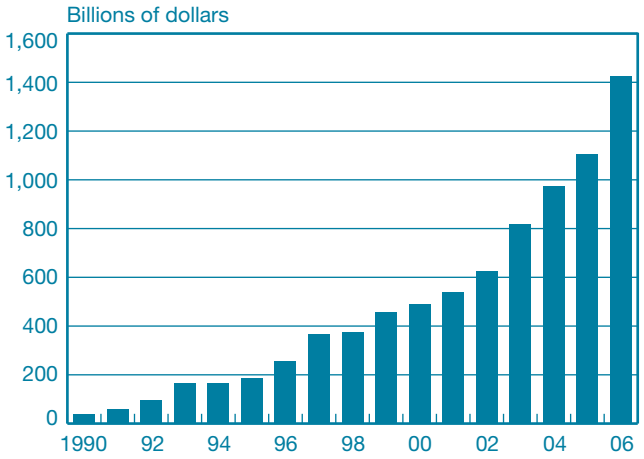
2.3 Industry Overview

By the end of 2006, the global hedge fund industry had about $1.43 trillion in assets[13] under management spread across more than 11,000 funds, one-third of which are fund of funds (European Central Bank 2006). However, because hedge funds are not required to register with any financial regulator or supervisor, these numbers can only be estimated. As we illustrate in the chart, the industry has grown enormously: in 1990, hedge funds had less than $400 billion in assets under management, and the $1 trillion mark was passed in 2005. Moreover, hedge funds are grabbing an increasing share of investable assets compared with mutual funds: in 1993, hedge fund assets under management were less than 4 percent of

[12]Mercer Oliver Wyman (2006); the sample of broker-dealers constituted 90 percent of the global business banks do with hedge funds.

[13]Source: Hedge Fund Research, Inc.

Total Assets under Global Management of Hedge Funds



Source: Hedge Fund Research, Inc.

mutual fund assets; that share increased to more than 10 percent by 2005 (Stulz 2007).

Hedge funds are also dominant players in several markets: in 2005, by one estimate, they accounted for 89 percent of U.S. trading volume in convertible bonds, 66 percent of volume in distressed debt, 33 percent of volume in emerging market bonds and in leveraged loans, 20 percent of speculative-grade-bond volume, and 38 percent of credit derivatives volume.[14] By early 2006, their estimated share of credit derivatives trading had increased to 58 percent (Greenwich Associates 2006). As these figures suggest, hedge funds are now engaged in a broader range of activity than in the past, especially in the trading of credit instruments.

Hedge funds come and go. Estimates of hedge fund survival rates vary between 85 percent and 95 percent (an attrition rate of 5 percent to 15 percent) per year, depending on the year and the style of fund. In their literature review, Chan et al. (2006) report that 30 percent of funds do not make it past three years, and 40 percent of funds do not survive past the fifth year. These survival rates are much lower than in the mutual fund industry, where average one-year attrition is less than 4 percent (Carhart et al. 2002). Although attrition rates are high for hedge funds, death by undercapitalization does not seem to be the main reason. Gupta and Liang (2005) report that nearly 90 percent of dead funds in their study were adequately capitalized at the time of closure.

[14]Source: Henny Sender and Anita Raghavan, "Private Money: The New Financial Order," *Wall Street Journal*, July 27, 2006, p. A1.

3. Systemic Risk

3.1 Defining Systemic Risk

Amid the rapid growth and innovation in global capital markets, financial stability and systemic risk have emerged as top policy concerns around the world. The Reserve Bank of Australia, the Bank of England, the Bank of Japan, the Norwegian Norges Bank, the Bank of Spain, the Swedish Riksbank, the Swiss National Bank, the Financial Stability Forum, the European Central Bank, and the International Monetary Fund, to name a few, publish regular reports on global financial conditions and financial stability issues.

Systemic risk, however, is not always defined and remains somewhat nebulous, so it is useful to be precise about what we mean by the term. In their exhaustive survey, DeBandt and Hartmann (2002) describe a "systemic crisis" as occurring when a shock affects:

> "a considerable number of financial institutions or markets in a strong sense, thereby severely impairing the general well-functioning (of an important part) of the financial system. The well-functioning of the financial system relates to the effectiveness and efficiency with which savings are channeled into the real investments promising the highest returns" (p. 11).

They define "systemic risk" as the risk of experiencing a systemic event.

Bordo, Mizrach, and Schwartz (1998) offer a similar description of a systemic event as a situation where:

> "shocks to one part of the financial system lead to shocks elsewhere, in turn impinging on the stability of the real economy" (p. 31),

while the Counterparty Risk Management Policy Group II (2005) describes a financial shock with systemic consequences as one with:

> "major damage to the financial system *and* the real economy" (p. 5, emphasis in original).

In our view, an essential feature of systemic risk is the potential of financial shocks to lead to substantial, adverse effects on the *real* economy, for example, by causing a reduction in productive investment by reducing credit provision or destabilizing economic activity. Indeed, it is the transmission of financial events to the real economy that is the defining feature of a systemic crisis, and what distinguishes it

from a purely financial event.[15] As we discuss in more detail below, these real effects might occur if credit provision is interrupted through shocks to the banking sector or through capital market disruptions.

This view also seems to reflect the thinking around the intervention during the collapse of Long-Term Capital Management (LTCM) in the fall of 1998. Federal Reserve Bank of New York President William J. McDonough (1998), in congressional testimony after the LTCM collapse, stated:

> "there was a likelihood that a number of credit and interest rate markets would experience extreme price moves and possibly cease to function for a period of one or more days and maybe longer . . . . Most importantly, this would have led to further increases in the cost of capital to American businesses."

Presumably, the increase in the cost of capital would have led to a reduction in credit provision and real activity would have suffered. McDonough's testimony explicitly stated that financial losses associated with asset price declines or failed trading strategies were not enough to motivate an intervention; rather, the concern was for "other market participants—investors who had no dealings with Long-Term Capital." This type of impact, either through direct exposures of particular intermediaries or broader disruptions to financial markets, provides a useful framework for discussing systemic risk as an "externality," which is a classic rationale for government intervention.

An important point is that the optimal level of systemic risk is not zero.

An important point is that the optimal level of systemic risk is not zero. A regulator, in principle, could eliminate all systemic risk by imposing sufficiently stringent limits on leverage or balance sheet linkages, or by imposing severe operating restrictions on key financial intermediaries, but this would unduly curtail the efficient activities of the financial sector and would be suboptimal from a social perspective. Without a fully developed model of the benefits of financial markets and the costs (and origins) of systemic risk, we obviously cannot determine the optimal level of systemic risk here, but it should be stressed that concern should focus on "inefficient" systemic risk that exceeds the socially optimal level.[16] As we discuss below, the idea of inefficient systemic risk is closely linked to market failures in counterparty credit risk management.

3.2 The Real Effects of Financial Intermediation

If a meaningful definition of systemic crisis involves a potential impact on real economic outcomes, the next step is to identify precisely the linkages between the financial sector in general, and hedge funds more specifically, and the real economy. This connection has historically been made through the "special" role that financial intermediaries, particularly banks, play in resolving information problems in the provision of credit. We begin with a discussion of this channel, and then broaden our perspective to consider the role of hedge funds as a potential source of systemic risk.

The traditional explanation for why banks matter for real economic activity rests on the assumption that borrowers (firms and entrepreneurs) are risky, but banks have unique skills that allow them to effectively screen lending opportunities, ex ante before investing, and then to monitor borrowers ex post. As it is inefficient for each saver to do this individually, banks become the "delegated monitors" that produce the critical information to facilitate the efficient allocation of credit. If bank lending activity becomes disrupted, due to insolvency or capital shocks, for example, socially productive relationships are severed and critical information is destroyed. As a result, some viable investment projects go unfunded and economic activity is reduced; ample evidence shows that bank lending affects real

The traditional explanation for why banks matter for real economic activity rests on the assumption that borrowers . . . are risky, but banks have unique skills that allow them to effectively screen lending opportunities, ex ante before investing, and then to monitor borrowers ex post.

[15] Bordo, Mizrach, and Schwartz (1998) label events that only lead to asset price declines and wealth losses without impinging the real economy as "pseudo-systemic risk" and conclude that "wealth losses are not synonymous with real systemic risk" (p. 33). The Counterparty Risk Management Policy Group II (2005) makes a similar distinction between a financial "disturbance" and a financial "shock," where only the latter has broad effects on the real economy.

[16] A simple model with diminishing marginal returns to financial intermediation and increasing marginal costs, for example, generates an optimal level of systemic risk that is not zero.

outcomes.[17] Note that this type of pure information problem tends to reduce the supply of credit (credit rationing). Later, we discuss how agency problems may reverse this and generate excess risk-taking.

Bank lending, however, is not the only information-intensive form of credit provision, and other forms are rising in relative importance. According to the Federal Reserve's Flow of Funds Accounts of the United States, bank credit accounted for 39 percent of outstanding credit market instruments for nonfarm, nonfinancial corporations in 2005, down from 52 percent in 1985, which reflects the growing importance of alternative sources of credit such as corporate bonds and commercial paper. These capital market instruments also rely on specific knowledge about borrowers' creditworthiness, so a financial disruption in these markets could also limit the provision of credit and have real economic effects.[18]

Financial intermediaries in this environment may not fund as much credit directly on their balance sheet, but act as underwriters, originators, and distributors of credit, and also as traders in the secondary market. Hedge funds, in particular, are active traders and contribute to increased market efficiency and liquidity through their frequent trading and ability to exploit arbitrage opportunities. This activity is generally stabilizing and provides considerable benefits in terms of greater market liquidity, lower volatility, and more stable relationships in the relative prices of financial assets. This should promote an efficient allocation of capital and improve real outcomes.

3.3 Hedge Funds and Systemic Risk

How, then, might hedge funds generate systemic risk? If systemic risk is fundamentally about financial market linkages to the real economy, then hedge funds create systemic risk to the extent that they can disrupt the ability of financial intermediaries or financial markets to efficiently provide credit. The President's Working Group on Financial Markets (1999), for instance, identifies the potential for the liquidation of a highly leveraged institution to lead to volatility and sharp asset price declines that "heighten uncertainty about credit risk and disrupt the intermediation of credit . . . heighten the risk of a contraction in real economic activity" (p. 23). A hedge fund link to the real economy might occur through banks' direct exposures to hedge funds, disruptions to capital markets that hinder credit provision or allocation, or indirect effects as bank problems feed back into the broader financial markets. We consider each point in turn and end with a discussion of what makes hedge funds particularly prone to exacerbating financial shocks.[19]

Commercial banks and securities firms are directly linked to hedge funds through their counterparty exposures, for example, short-run financing for leveraged positions, prime brokerage activity, and trading counterparty exposures in over-the-counter and other markets. If a bank has a large exposure to a hedge fund that defaults or operates in markets where prices are falling rapidly, the bank's greater exposure to risk may reduce its ability or willingness to extend credit to worthy borrowers. Collateralizing the credit exposures may not be enough to mitigate the risk. A sudden decline in asset prices triggered, for example, by the unwinding of a highly leveraged hedge fund can reduce the value of that collateral, or generate liquidity risk and further price declines via variation margining as investors sell into the falling market to meet margin calls. Such declines in collateral values, if sharp enough, can cast doubt on the assumptions relied upon in stress testing and risk management, and cause dealers to become more risk averse in their credit decisions (see Box 2). Moreover, according to the "financial accelerator" model popularized by Bernanke and Gertler (1989), a fall in asset values may reduce collateral values and thus hinder the ability of firms to borrow, which amplifies the impact of the initial shock.

If systemic risk is fundamentally about financial market linkages to the real economy, then hedge funds create systemic risk to the extent that they can disrupt the ability of financial intermediaries or financial markets to efficiently provide credit.

To the extent that bank-dependent borrowers cannot access alternative sources of funding, investment and economic activity will be curtailed until new relationships are formed and information recreated. This mechanism parallels the concern

[17] Ashcraft (2005) shows that the FDIC-induced failure of healthy banks in the 1980s led to persistent declines in relative income levels. A bank failure could also have real, adverse consequences if depositors are not made whole (their wealth is reduced) or if credit-constrained depositors lose liquidity, although deposit insurance and prompt resolution of failures limit these effects.

[18] Santos (2006) shows the importance of information as a determinant of access to the corporate bond market, especially during recessions, when signals of firm quality may be noisy.

[19] See McCarthy (2006) and Hildebrand (2007) for discussions of the transmission mechanism from hedge funds to financial stability that focus on the direct link between an intermediary and a hedge fund counterparty.

Box 2

Hedge Funds and Liquidity Risk

Hedge funds are typically viewed as being liquidity providers in the capital markets, which helps creates market efficiency.[a] Recent reports, however, suggest that hedge funds are moving increasingly into less liquid markets, with structured credit and distressed debt at the top of the list (European Central Bank 2006; Chan et al. 2006). In the presence of leverage, the combination of relatively illiquid assets and short-term financing exposes the hedge fund to possibly significant liquidity risk.

Hedge funds themselves, of course, are well aware of the consequences of their moves into less liquid markets, as are their counterparties. Perhaps as a result, hedge funds are adopting longer lock-up periods on their investors' ability to withdraw funds, which gives fund managers added flexibility to ride out market fluctuations (European Central Bank 2006; Mercer Oliver Wyman 2006). Mercer Oliver Wyman also points out the increased use of gates and notice periods for investor redemption, as well as contractual changes on the part of broker-dealers to increase transparency on hedge funds' liquidity positions. All of these market mechanisms act to reduce liquidity risk for hedge funds.

Nonetheless, liquidity management remains an important concern because of the potential impact on market dynamics. Before discussing the concern, it is useful to be precise and distinguish two types of liquidity: "market liquidity" is the ability to trade without affecting market prices, while "funding liquidity" is the ability to acquire funding in the event of credit impairment or some other shock. While distinct, these types of liquidity interact in important ways. When funding liquidity is abundant, for example, traders are able to finance positions, trade in higher volume to smooth price shocks, and make markets more liquid. In contrast, weak market liquidity tends to increase volatility, which leads to variation margin and collateral calls that reduce funding liquidity. Market liquidity shocks strain a trader's ability to fund its positions, as additional funds (for instance, to meet variation margin calls) can be raised only by selling assets into a falling market.

This mutual dependence creates the potential for market instability.[b] Consider the trading losses from a price shock. If the losses are severe enough to seriously erode traders' capital, then risk management trading limits that are defined relative to capital would compel traders to reduce their trading, and market liquidity would decline. At the same time, the increased volatility that typically accompanies price declines can lead to higher initial margin and collateral calls, which raise the cost of maintaining trading positions and reduce funding liquidity (Brunnermeier and Pedersen 2006). If the shock is large enough, a financially constrained trader will be compelled to sell assets, further depressing prices in the market under stress or transmitting the price shock to other markets as positions in those markets are liquidated to meet cash demands. These sales depress prices even more, causing a further negative shock to trading positions and setting in motion additional asset sales and a downward spiral in asset prices.

A natural question is what market mechanisms or policy responses would halt or reverse a downward liquidity spiral or bring about a recovery from an illiquid equilibrium. If hedge funds and other traders, who normally smooth out market imbalances and liquidity shocks, are themselves weakened by losses brought on by unusually large market shocks, which investors would step in as buyers, and what policies would encourage or promote their stabilizing behavior in a crisis? This is fundamentally a question about the limits to arbitrage, and it highlights some of the trade-offs that accompany hedge funds' growing role in financial markets.

[a] The role of hedge funds as liquidity providers has been documented; for example, the Agarwal, Daniel, and Naik (2007) study of the convertible bond market and the Mitchell and Pulvino (2001) study of merger arbitrage demonstrate hedge funds' role as creators of liquidity.

[b] Following the 1987 stock market crash, a growing body of research has arisen on trading-driven positive feedback in asset prices and liquidity. Among the earlier papers are DeLong et al. (1990), Gennotte and Leland (1990), and Grossman (1988). Recent examples include Brunnermeier and Pedersen (2006) and Shin (2006). Kambhu (2006) provides empirical evidence on the relationship between funding liquidity and market liquidity in convergence trading in the interest rate swap market.

in the early 1990s that bank loan losses in commercial real estate and the need to raise capital combined to create a "credit crunch" that exacerbated the U.S. recession in 1990-91. The concern now, of course, is that worrisome bank exposure is to hedge funds and capital markets, rather than to commercial real estate.

Discussions of this type of direct linkage from hedge funds to real economic activity through the banking system are common (see, for example, President's Working Group on Financial Markets [1999], Financial Stability Forum [2000, 2007], Chan et al. [2006], Garbaravicius and Dierick [2005],

Bernanke [2006], McCarthy [2006], and Hildebrand [2007]). While the magnitude of this exposure remains unclear, the Bank for International Settlements estimates that banks' direct exposure to hedge funds has been growing proportionately with the hedge fund industry itself.[20] It should be noted, however, that banks' current exposures to hedge funds are heavily collateralized and that the Financial Stability Forum (2007, p. 12) estimates that both the current and potential exposures, net of collateral, of core firms to hedge funds are quite modest in the aggregate. Moreover, each bank has a clear self-interest to manage and mitigate the risk of these exposures, although, as we discuss in Section 4, market discipline through counterparty credit risk management is not a panacea.

A second mechanism operates if hedge fund difficulties disrupt broad financial market activity, interrupt the efficient functioning of the capital markets, and hinder the broader provision of credit. It was feared, for example, that the market disruption surrounding the collapse of LTCM might impair the functioning of the credit and interest rate markets, and thus impede the provision of credit (McDonough 1998). This disruption of capital markets fundamentally reflects the loss of confidence of investors and a reduced ability or willingness to bear risk through the provision of credit (Counterparty Risk Management Policy Group II 2005). To the extent that this truly reflects an issue of investor confidence and not underlying fundamentals, it suggests an opportunity for policy intervention to shift market participants from a "bad" to a "good" equilibrium.

We emphasize that it is precisely the defining characteristics of hedge funds discussed earlier that create the potential for a substantial market disruption. For example, the funds' opacity and incentive structure may increase the likelihood of such an event as managers turn toward high-risk strategies with substantial tail-risk. Leverage, in turn, may amplify the impact of a given shock and result in larger and wider losses, as in the classic example of LTCM in the fall of 1998. Indeed, the President's Working Group on Financial Markets (1999), Chan et al. (2006), and others have highlighted excessive leverage as the key issue driving systemic concerns associated with hedge funds. Finally, the complexity and heterogeneity of the instruments may make the unwinding of positions more difficult, which would impede a timely and efficient workout and exacerbate the market impact.

A third systemic mechanism operates indirectly through the banking system. As discussed earlier, large commercial banks and broker-dealers provide substantial liquidity to the hedge fund sector by absorbing the counterparty credit exposure of trading positions, collateralizing financing, providing contingent credit lines, and making direct equity stakes. A hedge-fund-induced shock to a commercial bank could have knock-on effects if that bank (or other banks) reduces the provision of liquidity to other hedge funds or to other banks, and thus further disrupts financial markets and credit provision. Shin (2006), for example, argues that interlinkages of bank balance sheets create complex dynamics that can amplify an initial price shock.

These mechanisms are all conceptual in nature, and there is considerable uncertainty about how they might work in practice. For example, recent evidence shows that commercial banks provided considerable stabilization during the market disruption in the fall of 1998 (Saidenberg and Strahan 1999; Gatev, Schuermann, and Strahan 2006). In particular, when credit spreads rose and commercial paper markets dried up in October 1998, banks were in the position to provide credit, primarily through unused loan commitments and draw-downs on existing lines of credit, as transaction deposits flowed into the banking sector. Fearing that "growing caution by lenders and unsettled conditions in financial markets more generally [were] likely to be restraining aggregate demand in the future," the Federal Reserve decreased the target fed funds rate by 25 basis points on September 29, 1998, again on October 15, and once more on November 17, where it remained for a year.[21] In that episode, the Fed's injection of liquidity combined with normal market mechanisms alleviated the pressure from the short-run disruptions to the capital markets, so that a financial market crisis, driven by a hedge fund collapse, had minimal real effects and did not reach systemic proportions. When a crisis occurs, U.S. investors now seem to run to banks, not away from them. It is unclear, however, how effective this substitution of bank credit for capital market credit would be over longer periods if capital market disruptions persisted.

When a crisis occurs, U.S. investors now seem to run to banks, not away from them. It is unclear, however, how effective this substitution of bank credit for capital market credit would be over longer periods if capital market disruptions persisted.

[20] As reported in Garbaravicius and Dierick (2005).

[21] The announcement can be found at <http://www.federalreserve.gov/boarddocs/press/monetary/1998>.

4. Limitations to Counterparty Credit Risk Management

We now turn to our central question: Is CCRM, particularly by banks and securities firms, sufficient to limit risk-taking by hedge funds and constrain systemic risk to socially efficient levels? Financial regulators in the United States and abroad have for many years been guided by the principle that CCRM—not hedge fund regulation—is the optimal way to control hedge fund leverage and limit systemic vulnerabilities (see, for example, President's Working Group on Financial Markets [1999], Financial Stability Forum [2000, 2007], Counterparty Risk Management Policy Group II [2005], Financial Services Authority [2005], Parkinson [2006], and Bernanke [2006]). Most recently, the President's Working Group on Financial Markets (2007) concluded that "market discipline most effectively addresses the systemic risks posed by private pools of capital."

Financial regulators in the United States and abroad have for many years been guided by the principle that CCRM—not hedge fund regulation—is the optimal way to control hedge fund leverage and limit systemic vulnerabilities.

To assess the question of why CCRM might prove insufficient, it is useful to examine potential market failures (in a textbook sense of deviations from a perfectly competitive, full-information economy that efficiently allocates resources) in the provision of credit. These market failures include agency problems, externalities, free-rider problems, moral hazard, and coordination failures. We emphasize that these concerns apply more generally to many types of credit provision, but are likely more acute where information problems are most severe, where banks are eager to capture a share of a growing market, and where potential profits are encouraging stiff competition. Hedge fund exposures fit this description quite well, which makes them particularly vulnerable to the erosion of CCRM.[22]

[22]See Lacker (2006) for a skeptical perspective on the importance of market failures.

4.1 Agency Problems

An agency problem exists when participants have different incentives, and information problems prevent one party (the principal) from perfectly observing and controlling the actions of the second (the agent).[23] In this case, the agent may act in its own self-interest in a way that is detrimental to the principal. In terms of hedge funds, these agency problems may exist within the dealer/bank (for example, a trader versus a risk manager), within the hedge fund (for example, the hedge fund manager versus an investor), or in the credit relationship between the bank and the hedge fund.

An agency problem is likely to develop within the dealer/bank as a struggle between insiders (those who do the trading or establish the prime broker relationships) and outsiders (those who do not, such as risk managers, owners, or outside creditors). Because of different incentives and internal informational asymmetries, a trader or salesperson at a large dealer/bank may have less risk aversion and shorter horizons than the firm's management as his participation in the short-run upside exceeds participation in the downside (Allen 2003).[24] The opacity of the hedge fund counterparty is likely to exacerbate these difficulties, as it is harder for outsiders who are less informed of the fund's risk profile to determine the appropriate counterparty risk ratings that drive credit terms, such as initial margin or limits on the size of business conducted with the fund. These concerns are particularly severe when the normal business practice is to earn fees up-front, while potential losses fall in the future.[25] As a result, insiders' incentives will lead to excess risk-taking (from the firm's and society's perspective), which is possible in situations where information asymmetries prevent outsiders from perfectly observing, understanding, and controlling the actions of insiders (John and John 2006).

A second type of agency problem emerges between hedge fund managers and the fund's investors due to the combination

[23]This is a specific type of moral hazard, a situation where individuals maximize their own outcomes at the expense of others. This can occur when they do not bear the full consequences of their actions (a lack of insurance) or when information asymmetries prevent complete contracting (shirking on the job).

[24]Recent examples include a trader at Barings and one at Allied Irish Bank, both of whom made unauthorized trades that were subsequently hidden from the relevant risk managers (Allen 2003).

[25]This issue of compensation and risk-taking incentives is familiar to bankers who extend long-term loans when performance, including default, may not be known for many years. Indeed, a 1995 Federal Reserve Letter on Bank Lending Terms and Standards (Division of Banking Supervision and Regulation, SR95-36) addresses precisely this point with regard to corporate lending (see <http://www.federalreserve.gov/boarddocs/SRLETTERS/1995/sr9536.htm>).

of opacity of hedge fund strategies and the total-return-based compensation structures of hedge fund managers. Rajan (2005) warns about compensation structures in financial services that are more commonly convex, that is, increase strongly with good performance, but fall only mildly with poor performance. This convexity creates strong incentives to take on risks, and with lower transparency it is easier to hide those risks from both investors and counterparties. Moreover, Rajan points out that the risks that are more easily concealed "are 'tail' risks—that is, risks that have a small probability of generating severe adverse consequences and, in exchange, offer generous compensation the rest of the time." One such example where hedge funds are significant players is credit derivatives.[26] Rajan (2005) adds that:

> "[m]ost of the time, I will look as if I am outperforming my comparison group for I will have generated returns with no apparent risk. But every once in a while, disaster will strike and the creditor will default. My true risk profile will then be revealed but too late for my investors."

One might expect that the long-run self-interest of hedge fund managers would curtail excess risk-taking, but there is abundant anecdotal evidence that even managers of failed hedge funds are able to raise capital in subsequent funds. Similarly, traders who lose their jobs over trading losses are often able to obtain new employment as traders. The reason for these perhaps surprising facts is the inherent difficulty in distinguishing managers/traders with talent from those without talent, that is, to estimate "alpha," excess returns uncorrelated with market risk.

As a related point, there is a common perception that the convex payoffs from the high-water-mark contract of hedge fund managers increase incentives for risk-taking and contribute to agency problems. Rosenberg (2006) finds empirical evidence that the volatility of hedge funds' returns tends to increase as the returns fall below their high-water mark, and that the volatility increase is largest for funds with an incentive fee and high-water-mark provision. Panageas and Westerfield (2006), however, show that a high-water mark need not lead to greater risk-taking ex ante if hedge fund managers have a long horizon. Intuitively, the manager takes into account that being "under water" in the future is costly, increasing his risk aversion today.

In general, any solution to agency problems must balance the proper incentives with appropriate controls that limit excess risk-taking, such as credit risk mitigation practices that attempt to control exposure and loss given default. Practices such as initial and variation margin, collateralization of exposures, trading limits, and internal reporting systems can all serve this purpose. These practices are commonly used by dealers, but the critical question is how effectively they are implemented and under what conditions they may be waived. For example, a firm's risk manager's preferred level of initial margin to be imposed on a hedge fund may be quite different from the level preferred by the trader who owns an implicit option on the income generated by the trading relationship with the hedge fund.

[26] John and John (2006) show how such convex compensation structures can indeed increase systemic instability.

4.2 Externalities

An externality is an impact of one party's action on others who are not directly involved in the transaction. Credit exposure to hedge funds may create externalities in the banking system or broader financial markets in several ways. If the potential exposure amounts to a significant share of bank capital, for example, then a large shock to hedge funds could weaken banks and impair their ability to provide liquidity to the financial system or credit to borrowers. This can be considered an externality, as the impact is felt by market participants not directly involved in the original transaction, such as a corporate borrower that relies on a bank that suddenly becomes weakened.

A second way in which hedge fund exposures may generate externalities is in the common exposures to market risk factors across hedge funds and the dealer's own proprietary trading activity. The price impact of hedge funds' defensive trading or dynamic hedging after a shock to a position, for example, could further adversely affect the dealer's similar proprietary trading position. This, of course, relies on similarity in the market risk profiles of hedge funds and dealers, which could be aligned or offsetting. To the extent they are aligned, such a shock would potentially magnify the impact beyond the direct effect of the bank's credit exposure to hedge funds.

An externality is an impact of one party's action on others who are not directly involved in the transaction. Credit exposure to hedge funds may create externalities in the banking system or broader financial markets in several ways.

Whether these are in fact systemic issues depends on the size of the exposures relative to bank capital and the impact on the provision of credit. For large diversified banks, trading risk tends to be small relative to the credit risk in their loan book and small relative to current capital levels. With regard to credit exposure to hedge funds, the Financial Stability Forum (2007), for example, reports that the potential counterparty exposure of core firms to hedge funds is approximately between 3 percent and 10 percent of Tier 1 capital, and it concludes that "the size of direct exposure would not be alarming even if one assumed a wide margin of error" (p. 12).

Jorion (2006) examines the correlation of market risk capital charges and trading revenues across eleven large U.S. banks. Measured in this way, these banks turn out to have a fairly diversified exposure profile. O'Brien and Berkowitz (2006), using proprietary trading profit-and-loss data, show significant exposure heterogeneity across dealers. Another way to address the question of whether a correlated shock in the hedge fund and trading businesses is material relative to bank capital would be to conduct an integrated stress test across counterparty exposures and trading exposures at a dealer. A challenge in conducting such a stress test would be the selection of risk factors against which to assess the exposures. Many hedge funds and banks do not take simple directional trades against broad market risk factors. Instead, their trades tend to be spread trades between related securities, and these exposures would not be visible unless a stress test were conducted at a very high level of granularity in the representation of positions and risk factors. (See Box 3 for details on correlation risk and hedge funds.)

A related idea is the "public good" nature of financial market stability that generates a free-rider problem in terms of CCRM.[27] Consider, for example, a large hedge fund that has exposures with many banks, all of whom benefit from the health of the hedge fund. While in principle every bank should monitor its exposure and limit excess risk-taking by the hedge fund, each bank also has an incentive to free-ride by reducing its CCRM and enjoying the benefits of the CCRM of the other banks. This is a classic example of "tragedy of the commons," where private markets may underprovide the public good and create a rationale for official sector intervention.

This public-good concern may be exacerbated by liquidity risk. If very large positions are fundamentally different than smaller ones—for example, due to price feedback effects if the position needs to be liquidated in a time of stress, as described in Box 2—then CCRM may not provide enough collective discipline even in the absence of free-riding. This concern is amplified by the ability of a particular hedge fund to boost leverage through interaction with many dealers. While each dealer may have an incentive to monitor and control its own exposure, the opacity of a hedge fund's risk profile means that no particular firm would have the ability to control a fund's overall

[27] A public good is both nonrival (use by one person does not preclude use by others) and nonexcludable (available to all). It is related to an externality because one person's provision of the good benefits others.

Box 3
Hedge Funds and Correlations

Are hedge fund returns more correlated than the returns of other asset managers? And if so, does the increased correlation stem from exposure to common risk factors or some other source of dependence or commonality? Finally, do these correlations get "worse" during periods of market turmoil? Affirmative answers to these questions would cause concern about the unusual effect hedge funds may have on the financial system in adverse market conditions.

As perhaps befitting their name, hedge fund returns tend to be only weakly (linearly) correlated with broad market returns (Fung and Hsieh 1999; Boyson, Stahel, and Stulz 2006), although Garbaravicius and Dierick (2005) provide evidence that this correlation has been increasing in recent years. Yet the European Central Bank (2006, p. 134) documents that hedge funds have become more correlated with each other: "the levels reached in late 2005 exceeded those that had prevailed just before the near-collapse of [Long-Term Capital Management]." Adrian (2007), however, concludes that this primarily reflects lower overall volatility in the recent period.

The picture becomes somewhat more complex when one looks beyond linear dependence. Boyson, Stahel, and Stulz (2006) analyze co-occurrences of extreme moves (5 percent to 15 percent tail events), which they call contagion. They find no supporting evidence of contagion from extreme moves in the underlying risk factors—equity, fixed-income, and currency markets—but they do find evidence of contagion across hedge fund styles. One interpretation is that this simply reflects omitted risk factors, such as liquidity risk, which are common across styles. Alternatively, these findings may lend support to Rajan's (2005) assertion of herding behavior among asset managers generally, a concern also raised by the Counterparty Risk Management Policy Group II (2005) in what it calls "crowded trades," but disputed by EDHEC Risk and Asset Management Research Centre (2006). Overall, however, the conclusion drawn by Boyson, Stahel, and Stulz (2006) is worth noting:

> "These results imply that systemic risk arising from poor performance in the broad markets is not amplified by contagion to the hedge fund industry" (p. 31).

leverage and exposure to risk. In this situation, the competitive equilibrium will not impose enough collective discipline.

4.3 Moral Hazard

Moral hazard refers to changes in behavior in response to redistribution of risk, for example, insurance may induce risk-taking behavior if the insured does not bear the full consequences of bad outcomes. In financial markets, the question of moral hazard from conjectural guarantees by the government—the implicit promise to bail out certain bank creditors—may apply to the largest commercial banks, but it does not apply to hedge funds. The resolution of the LTCM crisis, for example, was essentially an informal bankruptcy procedure in which LTCM's stakeholders were largely wiped out in the recapitalization and ultimate liquidation of the fund by its dealer counterparties. Further, regulators continually disavow any "too-big-to-fail" policy for hedge funds.

In terms of the regulated largest financial intermediaries, however, their central role in the financial system raises the question of how large credit losses would affect regulators' decisions about the regulated firm. As a result, a bank (and its counterparties) may arguably have less incentive to monitor, reduce risk, and limit exposure in activities whose current profitability and growth opportunities make them attractive to banks—such as in the hedge fund sector. This moral hazard issue for large financial intermediaries is relevant in all of their risk management decisions, however, not just those related to their hedge fund business. Furthermore, the substantial franchise value attributable to large financial institutions should help to mitigate somewhat these moral hazard issues.

4.4 Competition

A final concern about the breakdown of effective CCRM is that the apparent profits to be earned in this business may create competitive pressures that weaken credit risk mitigation practices. Bernanke (2006) and the Financial Stability Forum (2007), for example, discuss how competition for new hedge fund business may be eroding CCRM, such as through lower than appropriate fees and spreads, or inadequate risk controls such as lower initial margin levels, collateralization practices, or exposure limits.[28]

[28] This is not a new concern. See President's Working Group on Financial Markets (1999) for similar concerns about potentially harmful effects of competition for hedge fund business.

While this is not a market failure per se, the U.S. banking sector has a history of periods in which lax CCRM contributed to substantial credit losses and potential systemic consequences: the less-developed-country debt crisis of the late 1970s and early 1980s; excessive commercial real estate lending in the late 1980s; possibly the weak counterparty credit measures in the 1990s that allowed LTCM to take on enormous leverage in many markets; and most recently the subprime mortgage market. As mentioned earlier, this deterioration may partially reflect inefficient compensation schemes and short horizons of lenders, although some portion of the adverse outcomes simply reflects the underlying, negative shocks.

A complicating factor, however, is that competition and expected profit typically improve efficiency and erode economic rents by inducing entry. This makes it more difficult for outside observers to assess whether claims of competitive excess are real or the complaint of incumbents whose profits are being eroded by competition. This type of competition may actually be socially efficient if earlier margins, for example, were "too high" and unduly restricted activity.

A final concern about the breakdown of effective CCRM is that the apparent profits to be earned in this business may create competitive pressures that weaken credit risk mitigation practices.

The economics literature has revealed very little on the interaction of competition and risk management discipline on practices such as initial margin levels. This is because relevant data are not available, for example, on the distribution of dealers' hedge fund business by counterparty risk class, how it has changed, how measures of hedge fund creditworthiness vary across dealers, and how they relate to potential future exposure and risk mitigation practices such as initial margin. This remains an important question for future research.

5. Implications for Regulation

A typical economist's view is that market participants enter into transactions with a full understanding of the benefits and costs of their actions. To the extent that participants are well-informed and all the costs and benefits accrue to those making

the transaction, a perfectly competitive environment leads to a laissez-faire level of systemic risk that is socially efficient, so there is no rationale for regulation or intervention. Additional direct regulation would likely reduce social welfare due to less activity, entry deterrence, stifled innovation, limited competition, distorted behavior and regulatory arbitrage, increased moral hazard, and direct compliance costs.

As we discussed earlier, however, the textbook description of perfect competition is typically not appropriate in all financial and credit markets, where agency problems, externalities, and moral hazard are common. As a result, the laissez-faire level of systemic risk may be too high from society's perspective, raising the question of whether regulation may potentially improve outcomes. Indeed, a key lesson from the collapse of LTCM is that market participants may not be sufficiently cognizant of the risks they face and therefore not vigilant enough in constraining counterparty risk.

A key lesson from the collapse of [Long-Term Capital Management] is that market participants may not be sufficiently cognizant of the risks they face and therefore not vigilant enough in constraining counterparty risk. Since LTCM, however, CCRM has greatly improved.

Since LTCM, however, CCRM has greatly improved. The Counterparty Risk Management Policy Group II (2005), Mercer Oliver Wyman (2006), and the Financial Stability Forum (2007), for example, note improved risk management techniques by counterparties, improved supervision, more effective disclosure and transparency, strengthened financial infrastructure, and more effective hedging and risk distribution techniques. Moreover, the institutionalization of hedge funds, driven by demands of new investors such as hedge fund of funds and other institutional investors, is increasing discipline and transparency. More research on the role of hedge fund of funds as both a stabilizing force (through increased discipline and reallocation of capital to better performers) and a destabilizing force (as a source of "hot money") would likely be fruitful.

Despites these gains, if systemic risk reflects an externality or public-good problem, then by definition even well-informed market participants will not have an incentive to adequately monitor or limit those risk-generating actions and there is a role for regulation to reduce inefficient systemic risk (President's Working Group on Financial Markets 1999; Bernanke 2006). Effective policy, however, needs to be able to precisely identify and to quantify the externality in order to determine the appropriate policy (Financial Services Authority 2005). Moreover, we have no evidence that externalities relating to hedge funds are unique. Therefore, a second area of research is to examine whether externalities relating to hedge funds are more acute than in other areas of bank lending, where stronger CCRM remains the appropriate policy response.

A more forceful alternative—outright regulation of hedge funds such as through activity restrictions, required capital, or leverage restrictions—has not received much attention and could have substantial costs. Activity restrictions that dramatically limit trading strategies such as short-selling or the use of derivative investments, for example, would likely diminish the beneficial effect of hedge funds on market liquidity and price discovery, thereby reducing the benefits along with the costs. Required capital ratios would be difficult to set optimally and would likely lead to increased regulatory arbitrage. Outright regulation might be expected to increase moral hazard if it increases the appearance of regulatory approval or simply the shift of activity to a less regulated jurisdiction. With a heavy regulatory hand, there is a risk of hedge funds moving totally off-shore; regulators might go from seeing little to seeing nothing. Finally, the historical policy response to lax CCRM has not been to regulate the borrower, but to increase oversight of the lenders, so regulation of hedge funds would be a significant departure from policy precedents.[29]

A second alternative is the mandatory provision of more information to regulators and the investment community. In principle, better informed investors would be increasingly able to monitor and discipline hedge funds, and thus reduce excess risk-taking. One specific idea raised by the President's Working Group on Financial Markets (1999) was improvement in the disclosure of hedge funds' risk profiles to their investors and counterparties. Timely and meaningful disclosures that do not compromise legitimate commercial interests of a hedge fund would help address the information asymmetries related to the market failures discussed above. Another specific idea raised by some is the creation of a large database, maintained by the official sector, that aggregates and records hedge fund

[29]In response to excessive commercial real estate lending in the property boom of the 1980s, regulators established stronger supervisory guidance on property lending, while the banks and property developers devised innovative financing arrangements involving securitization that shifted some of the risk out of the banking sector. Significantly, no one proposed that commercial property developers become regulated institutions.

exposures on a real-time basis. It is not clear, however, how the official sector could effectively analyze the enormous quantity of information or act on it given the heterogeneous nature of positions and exposures, and this would undoubtedly be a very costly exercise (Bernanke 2006). Moreover, disclosure does not solve the externality problem or address all of the agency issues.

A third alternative, discussed by the Financial Services Authority (2005) and the Counterparty Risk Management Policy Group II (2005), is to encourage through moral suasion "best-practice" techniques for risk management and measurement, both within institutions and in terms of market infrastructure. Meaningful best-practice efforts spearheaded by the official sector or industry groups can leverage market discipline and encourage institutions to meet generally agreed-upon standards in terms of accounting, transparency, and risk management. For market infrastructure issues, the recent work led by the Federal Reserve Bank of New York and other regulators on credit derivatives clearing and settlement shows how market participants may not have adequate private incentives to provide essential "plumbing" that is good for all. This is a classic example of a public good that needs to be provided by the official sector to avoid the "tragedy of the commons," in this case, systemic risk. Other examples may include improved clarity around legal arrangements for prime brokerage activities and valuation standards for complex, illiquid products.

We conclude that the current emphasis on market discipline and CCRM as the primary check on hedge fund risk-taking is appropriate. If systemic risk were to originate through direct banking sector exposures, for example, then the banks themselves would have the strongest incentive to monitor those exposures and limit risk. While various market failures may make CCRM imperfect, it remains the best line of defense against systemic risk.

References

Ackermann, C., R. McEnally, and D. Ravenscraft. 1999. "The Performance of Hedge Funds: Risk, Returns, and Incentives." Journal of Finance 54, no. 3 (June): 833-74.

Adrian, T. 2007. "Measuring Risk in the Hedge Fund Sector." Federal Reserve Bank of New York Current Issues in Economics and Finance 13, no. 3 (March/April).

Agarwal, V., N. D. Daniel, and N. Y. Naik. 2007. "Role of Managerial Incentives and Discretion in Hedge Fund Performance." Unpublished paper, Georgia State University, March 1. Available at <http://ssrn.com/abstract=889008>.

Allen, S. 2003. Financial Risk Management: A Practitioner's Guide to Managing Market and Credit Risk. Hoboken, N.J.: John Wiley and Sons.

Almazan, A., K. C. Brown, M. Carlson, and D. A. Chapman. 2004. "Why Constrain Your Mutual Fund Manager?" Journal of Financial Economics 73, no. 2 (August): 289-321.

Ashcraft, A. 2005. "Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks." American Economic Review 95, no. 5 (December): 1712-30.

Banque de France. 2007. "Special Issue on Hedge Funds." Financial Stability Review 10, April. Available at <http://www.banque-france.fr/gb/publications/rsf/rsf_042007.htm>.

Bernanke, B. S. 2006. "Hedge Funds and Systemic Risk." Remarks delivered at the Federal Reserve Bank of Atlanta's 2006 Financial Markets Conference—Hedge Funds: Creators of Risk? May 16.

Bernanke, B. S., and M. Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." American Economic Review 79, no. 1 (March): 14-31.

Bordo, M. D., B. Mizrach, and A. J. Schwartz. 1998. "Real versus Pseudo-International Systemic Risk: Some Lessons from History." Review of Pacific Basin Financial Markets and Policies 1, no. 1 (March): 31-58.

Boyson, N. M., C. W. Stahel, and R. M. Stulz. 2006. "Is There Hedge Fund Contagion?" NBER Working Paper no. 12090, March.

Brunnermeier, M., and L. Pedersen. 2006. "Market Liquidity and Funding Liquidity." Unpublished paper, Princeton University.

Carhart, M. M., J. N. Carpenter, A. W. Lynch, and D. K. Musto. 2002. "Mutual Fund Survivorship." Review of Financial Studies 15, no. 5 (winter): 1439-63.

Chan, N., M. Getmansky, S. M. Haas, and A. W. Lo. 2006. "Systemic Risk and Hedge Funds." In M. Carey and R. Stulz, eds., Risks of Financial Institutions, 235-330. Chicago: University of Chicago Press.

Counterparty Risk Management Policy Group II. 2005. "Toward Greater Financial Stability: A Private Sector Perspective." July 27. Available at <http://www.crmpolicygroup.org/>.

DeBandt, O., and P. Hartman. 2002. "Systemic Risk: A Survey." In C. A. E. Goodhard and G. Illing, eds., Financial Crisis, Contagion, and the Lender of Last Resort: A Book of Readings. London: Oxford University Press.

DeLong, J. B., A. Schleifer, L. Summers, and R. Waldman. 1990. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." Journal of Finance 45, no. 2 (June): 379-95.

Demsetz, R. S., and P. E. Strahan. 1997. "Diversification, Size, and Risk at Bank Holding Companies." Journal of Money, Credit, and Banking 29, no. 3 (August): 300-13.

EDHEC Risk and Asset Management Research Centre. 2006. "Disorderly Exits from Crowded Trades? On the Systemic Risks of Hedge Funds." Available at <http://www.edhec-risk.com/edito/RISKArticleEdito.2006-07-27.4050>. Registration required.

Elsinger, H., A. Lehar, and M. Summer. 2006. "Risk Assessment for Banking Systems." Management Science 52, September: 1301-14.

Elton, E. J., M. J. Gruber, and C. R. Blake. 2003. "Incentive Fees and Mutual Funds." Journal of Finance 58, no. 2 (April): 779-804.

European Central Bank. 2006. Financial Stability Review. June. Available at <http://www.ecb.int/pub/pdf/other/financialstabilityreview200606en.pdf>.

References (Continued)

Financial Services Authority. 2005. "Hedge Funds: A Discussion of Risk and Regulatory Engagement." Discussion Paper no. 05/4, June.

Financial Stability Forum. 2000. "Report of the Working Group on Highly Leveraged Institutions." April 5.

———. 2007. "Update of the FSF Report on Highly Leveraged Institutions." May 19.

Flannery, M. J., S. H. Kwan, and M. Nimalendran. 2004. "Market Evidence on the Opaqueness of Banking Firms' Assets." Journal of Financial Economics 71, no. 3 (March): 419-60.

Fung, W., and D. A. Hsieh. 1999. "A Primer on Hedge Funds." Journal of Empirical Finance 6, no. 3 (September): 309-31.

Fung, W., D. Hsieh, N. Naik, and T. Ramadorai. 2006. "Hedge Funds: Performance, Risk, and Capital Formation." Unpublished paper.

Garbaravicius, T., and F. Dierick. 2005. "Hedge Funds and Their Implications for Financial Stability." European Central Bank Occasional Papers, no. 34, August.

Gatev, E., T. Schuermann, and P. E. Strahan. 2006. "How Do Banks Manage Liquidity Risk? Evidence from the Equity and Deposit Markets in the Fall of 1998." In M. Carey and R. Stulz, eds., Risks of Financial Institutions, 105-30. Chicago: University of Chicago Press.

Gennotte, G., and H. Leland. 1990. "Market Liquidity, Hedging, and Crashes." American Economic Review 80, no. 5 (December): 999-1021.

Greenwich Associates. 2006. "For Hedge Funds, Fixed-Income Trading Volumes Soar, While Costs Take on New Importance." Research report, August.

Grossman, S. 1988. "An Analysis of the Implications for Stock and Futures Price Volatility of Program Hedging and Dynamic Hedging Strategies." Journal of Business 61, no. 3 (July): 275-98.

Gupta, A., and B. Liang. 2005. "Do Hedge Funds Have Enough Capital? A Value-at-Risk Approach." Journal of Financial Economics 77, no. 1 (July): 219-53.

Hahn, R. W., and C. R. Sunstein. 2005. "The Precautionary Principle as a Basis for Decision Making." The Economists' Voice 2, no. 2 (Article 8).

Hildebrand, P. M. 2007. "Hedge Funds and Prime Broker-Dealers: Steps Towards a 'Best Practice Proposal.'" Banque de France Financial Stability Review 10, April: 73-83.

John, K., and T. A. John. 2006. "Managerial Incentives, Derivatives, and Stability." Journal of Financial Stability 2, no. 1 (April): 71-94.

Jorion, P. 2006. "Bank Trading Risk and Systemic Risk." In M. Carey and R. Stulz, eds., Risks of Financial Institutions, 29-58. Chicago: University of Chicago Press.

———. 2007. Value at Risk. 3rd ed. New York: McGraw Hill.

Kambhu, J. 2006. "Trading Risk, Market Liquidity, and Convergence Trading in the Interest Rate Swap Spread." Federal Reserve Bank of New York Economic Policy Review 12, no. 1 (May): 1-13.

Lacker, J. M. 2006. "How Should Regulators Respond to Financial Innovation?" Remarks delivered at The Philadelphia Fed Policy Forum 2006, December 1. Available at <http://www.philadelphiafed.org/econ/conf/forum2006/papers/lacker_speech.pdf>.

Lowenstein, R. 2000. When Genius Failed: The Rise and Fall of Long-Term Capital Management. New York: Random House.

McCarthy, C. 2006. "Hedge Funds: What Should Be the Regulatory Response?" Remarks delivered at the SUERF European Money and Finance Forum 2006, December 7. Available at <http://www.fsa.gov.uk/pages/Library/Communication/Speeches/2006/1207cm.shtm>.

McDonough, W. J. 1998. "Statement before the Committee on Banking and Financial Services." U.S. House of Representatives, October 1.

Mercer Oliver Wyman. 2006. "New Study Reveals Strengthened Global Hedge Fund Industry Risk Management Practices and Highlights Areas for Improvement." Press release, June 20. Available at <http://www.oliverwyman.com/ow/pdf_files/Global_Hedge_Fund_RM_MOW_0606.pdf>.

References (Continued)

Mitchell, M., and T. Pulvino. 2001. "Characteristics of Risk and Return in Risk Arbitrage." Journal of Finance 56, no. 6 (December): 2135-75.

Morgan, D. P. 2002. "Rating Banks: Risk and Uncertainty in an Opaque Industry." American Economic Review 92, no. 4 (September): 874-88.

O'Brien, J., and J. Berkowitz. 2006. "Estimating Bank Trading Risk: A Factor Model Approach." In M. Carey and R. Stulz, eds., Risks of Financial Institutions, 59-101. Chicago: University of Chicago Press.

Panageas, S., and M. Westerfield. 2006. "High-Water Marks: High-Risk Appetites? Convex Compensation, Long Horizons, and Portfolio Choice." Unpublished paper, University of Southern California and University of Pennsylvania, Wharton School.

Paredes, T. A. 2006. "On the Decision to Regulate Hedge Funds: The SEC's Regulatory Philosophy, Style, and Mission." Washington University in St. Louis School of Law, Faculty Working Paper Series, no. 6-03-02.

Parkinson, P. M. 2006. "The Role of Hedge Funds in the Capital Market." Testimony before the Subcommittee on Securities and Investment, Committee on Banking, Housing, and Urban Affairs. U.S. Senate, May 16.

President's Working Group on Financial Markets. 1999. "Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management." April 28. Available at <http://www.treas.gov/press/releases/reports/hedgfund.pdf>.

———. 2007. "Agreement among PWG and U.S. Agency Principals on Principles and Guidelines Regarding Private Pools of Capital." Available at <http://www.treas.gov/press/releases/reports/hp272_principles.pdf>.

Rajan, R. G. 2005. "The Greenspan Era: Lessons for the Future." Remarks delivered at the Symposium on Financial Markets, Financial Fragility, and Central Banking, sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 27.

Rosenberg, J. 2006. "When Hedge Funds Underperform: The Impact of Manager Compensation Type on Risk-Taking." Unpublished paper, Federal Reserve Bank of New York.

Saidenberg, M. R., and P. E. Strahan. 1999. "Are Banks Still Important for Financing Large Businesses?" Federal Reserve Bank of New York Current Issues in Economics and Finance 5, no. 12 (August).

Santos, J. A. C. 2006. "Why Firms' Access to the Bond Market Differs over the Business Cycle: A Theory and Some Evidence." Journal of Banking and Finance 30, no. 10 (October): 2715-36.

Securities and Exchange Commission. 2003. "Implications of the Growth of Hedge Funds." Staff report to the United States Securities and Exchange Commission, September.

Shin, H. S. 2006. "Risk and Liquidity in a System Context." BIS Working Paper no. 212, August.

Stulz, R. M. 2007. "Hedge Funds: Past, Present, and Future." Journal of Economic Perspectives 21, no. 2 (June): 175-94.

Robert Rich and Charles Steindel

A Comparison of Measures of Core Inflation

- Designed to remove transitory changes from aggregate price data, measures of underlying, or "core," inflation are important tools in the monetary policymaking process.
- Somewhat surprisingly, little consensus has been reached on a preferred measure of U.S. core inflation.
- An evaluation of several proposed measures of U.S. core inflation, including the popular ex food and energy series, finds that no measure consistently dominates the others.
- There is arguably too much variability in the nature and sources of transitory price movements to be captured effectively through the design of any individual measure.
- The general practice of focusing on a measure of core inflation that excludes food and energy does not seem to be justified by the analysis.

1. Introduction

Central banks differ in their specific inflation objectives and conduct of policy. However, they typically confront the common problem of differentiating which price changes are permanent and which are transitory. Because of the lagged effects of monetary policy, mistaking the nature of price changes can be extremely costly. For example, the failure to detect the onset of inflationary pressures may lead to a sustained rise in inflation and ultimately require a more prolonged period of policy tightening. Then again, an overreaction to a temporary increase in inflation may result in an unwarranted slowing, and possible decline, in economic activity. Thus, the ability of central banks to differentiate between permanent and transitory price movements is critical for determining the appropriate prescription for monetary policy.

The importance of gauging the persistence of price changes in a timely manner has led to the development of schemes to filter incoming data on aggregate prices. The filtering schemes are designed to remove transitory price movements and thereby produce a measure of underlying, or "core," inflation. The most common measure of core inflation is aggregate household inflation excluding the contribution of price

Robert Rich is a research officer and Charles Steindel a senior vice president at the Federal Reserve Bank of New York.
<robert.rich@ny.frb.org>
<charles.steindel@ny.frb.org>

The authors have benefited from the suggestions of attendees at the Bank for International Settlements conference on the Evolving Inflation Process and the Federal Reserve Bank of Cleveland–Federal Reserve Bank of Dallas conference on Price Measurement for Monetary Policy. They thank Adam Ashcraft, Steve Cecchetti, Tim Cogley, and two anonymous referees for valuable comments, and are grateful to Steven Reed of the Bureau of Labor Statistics for supplying detailed methodologically consistent CPI data. They also thank David Bedoll, Michael Strain, and Evan LeFlore for research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

changes from food and energy. However, alternative core inflation measures have been proposed. Some of these candidate series include ex energy measures (Clark 2001), weighted median measures (Bryan and Cecchetti 1994), and exponentially smoothed measures (Cogley 2002), with proponents citing the superior properties and performance of the respective series across various dimensions.

The lack of consensus on a preferred measure of core inflation might seem surprising given the importance of this information to policymakers. However, a closer examination of the evidence reveals little uniformity across the dimensions used to compare the proposed measures. These dimensions include statistical metrics, such as within-sample regression fit and out-of-sample forecast performance as well as more basic considerations relating to the selection of the sample period for the analysis; the data frequency of the price changes (that is, monthly versus quarterly observations); and even the choice of the price measure used to calibrate the core inflation series. Consequently, it is unclear whether the conflicting evidence in support of various measures of core inflation reflects inherent differences in performance capabilities or a lack of standardization in the evaluation process.

The lack of consensus on a preferred measure of core inflation might seem surprising given the importance of this information to policymakers. However, a closer examination of the evidence reveals little uniformity across the dimensions used to compare the proposed measures.

This article provides a systematic evaluation of several proposed measures of U.S. core inflation: the popular ex food and energy series, an ex energy series, a weighted median series, and an exponentially smoothed series. To inform the current debate on this issue, we adopt a general framework for the analysis. Regarding the choice of aggregate price indexes, we focus on inflation measures that are likely goals for U.S. monetary policy, namely the consumer price index (CPI) and the personal consumption expenditure (PCE) index. To evaluate the core measures of inflation, we select criteria that have been used in previous studies: ease of design, a similar mean to the goal inflation series, and an ability to track the trend in the goal inflation series.

We also include the explanatory and forecasting capabilities of the core measures of inflation as criteria, but recognize that the need to specify an econometric model introduces some discretion into the analysis. In an attempt to mitigate this concern, we adopt a benchmark model that relates future changes in inflation to the transitory component of price changes identified by the candidate series. This specification has the advantage of being not only simple to interpret in this context, but also flexible enough to allow us to incorporate alternative horizons into the analysis. As a further check for robustness, we also examine results over different subsamples.

Taken together, we find that no core measure of inflation consistently dominates the others. Further, the performance of the candidate series differs markedly across the aggregate inflation measures, criteria, and sample periods. This conclusion is unaffected by the addition of simple measures of economic slack to the benchmark model. Therefore, we contend that the specifics of the criteria and methods used likely do not account for the differences in performance capabilities of the candidate series. Rather, we suggest that this conflicting evidence reflects the lack of a consistent pattern in transitory price movements. Namely, there is too much variability in the nature and sources of transitory price movements to be captured effectively through the design of the individual core inflation measures. We argue that this interpretation is consistent with the diversity of previous findings using U.S. data and with the work of Hogan, Johnson, and Laflèche (2001) in the Canadian context and Mankikar and Paisley (2002) in the U.K. context. Both studies similarly conclude that there is no single core inflation measure that performs well across-the-board.

Our inability to identify a clear "best" or "worst" measure of either core CPI or core PCE inflation also has implications for some aspects of policy formulation and discussion. While it would be desirable to rely on a single measure of core inflation to perform a multitude of tasks, the evidence does not offer support for this scenario. Consequently, we cannot identify a compelling analytical reason, on either an ex ante or ex post basis, to concentrate attention on a measure of inflation that excludes food and energy prices.

2. Motivation and Concepts

Almost all central banks are concerned with, and have some mandate to achieve, price stability. Even when ongoing changes in the aggregate price level are anticipated, however, the changes impose costs on economies. These costs need not be directly related to movements in any type of household price measure; they could stem from systematic changes in the prices of all goods and services produced or purchased, including items bought by businesses and governments.

As a practical matter, inflation goals are often linked to movements in a price measure for goods and services purchased by consumers. One reason for this linkage is that the prices for many capital goods purchased by businesses are extremely difficult to measure,[1] as are those for many products provided by governments (such as public education). Moreover, a broad measure of consumer prices should be reasonably successful in capturing the component of aggregate price movements that may affect economic efficiency.

Another reason why a central bank would be concerned with movements in a household or consumer price measure is that many formal escalation arrangements, notably for wages as well as taxes and government benefits, are connected to indexes of consumer prices.[2] These arrangements could lead to household price movements affecting the distribution of income as well as government revenues and expenditures. In turn, these shifts could influence employment, investment, and basic fiscal policy decisions, and thereby affect the macroeconomic environment faced by monetary policymakers. Thus, there are pragmatic reasons for central banks to concentrate their attention on the consumer component of inflation.

Given a concern with longer term movements in household price inflation, central banks and private agents need some means by which to gauge current performance vis-à-vis a price inflation objective.[3] The main reason to focus on the behavior of a core price measure is the belief that there is a significant amount of transitory noise in the movement of aggregate consumer prices. Filtering out the transitory noise gives a better sense of the underlying trend in prices, and thus a better sense of how a measure of current price changes compares with an explicit or inferred longer term goal.[4] Accordingly, the role of a core price measure lends itself to being interpreted as a means to an end, with low and stable growth of a core price measure serving as an "intermediate target" of policy rather than as a direct "goal."[5] This interpretation might also make clear that a central bank's decision to downplay certain price changes in the conduct and communication of monetary policy does not indicate a lack of concern for the impact of these price changes on current movements in the cost of living.[6]

The development of the core inflation concept appears to have begun in the early 1970s. An early (and likely initial) construction, associated with the late Otto Eckstein, was a weighted growth of unit labor and capital costs for the economy as a whole (Eckstein 1981). The more familiar measure of core inflation as aggregate price growth excluding food and energy appears to have been analyzed first in a systematic fashion in a paper by Gordon (1975b). Gordon's aggregate "'core' price equation" was estimated for final sales prices excluding food and energy.[7]

The main reason to focus on the behavior of a core price measure is the belief that there is a significant amount of transitory noise in the movement of aggregate consumer prices.

The name "core inflation" then began to be attached to the growth of price measures excluding food and energy. In 1978, the Bureau of Labor Statistics began to report monthly growth of both the CPI and the producer price index excluding food and energy. The Bureau of Economic Analysis also releases data on the monthly growth of the PCE index excluding food and energy as well as the "market-based" PCE index excluding food and energy. An important point concerning the development of these "core inflation" measures is that little or no specific consideration was given for their future use in the formulation of monetary policy.

[1]One could argue that the ideal aggregate inflation index would not include the acquisition prices of capital goods, but rather would include the current "user cost" of existing capital. Nonetheless, as is the case for capital goods acquisition prices, it is difficult to measure these user costs accurately.

[2]For instance, increases in the U.S. CPI automatically increase federal income tax brackets and some deductions and exemptions as well as trigger boosts in social security benefits, federal employee pensions, and interest payments on inflation-protected securities.

[3]The issue of whether or not a price inflation objective should be stated as a numerical inflation target is not relevant to our analysis. Our focus is the construction of a measure of underlying inflation that both satisfies some given criteria and is useful for policymakers and private agents concerned with the ongoing path of price changes.

[4]As in Mankikar and Paisley (2002) and Brischetto and Richards (2006), one can alternatively describe the role of a core price measure in terms of distinguishing between relative price movements and changes in underlying inflation.

[5]Some researchers (Aoki 2001; Benigno 2004; Goodfriend and King 1997) have argued that the appropriate goal for monetary policy should be set in terms of a measure of "core" inflation. However, these authors are referring to a measure that comprises "sticky" prices—those prices that change at fixed intervals—and excludes "flexible" prices that may change at any time. The authors' use of this terminology may stem from the view that their goal inflation series is somewhat comparable to the aggregate index less food and energy, since food and energy prices may be much more flexible than others. Nevertheless, it is important to note that the models underlying this argument are highly stylized. Moreover, while the argument that policy should be concerned primarily with changes in "sticky" prices to offset the resulting inefficiencies may have merits, there is the more overriding concern that it is very difficult to develop an index of "stickiness" to evaluate candidate core inflation measures.

[6]That is, this treatment may alleviate the apparent disconnect and occasional sharp deviation between changes in a cost-of-living index and in the inflation measure(s) that may be the focus of central bank discussion.

[7]In a slightly earlier piece, Gordon (1975a) refers to 1973-74 inflation as comprising several components, including "underlying 'hard-core' inflation" (p. 184).

3. Core Inflation: Proposed Measures and Evaluation

3.1 Candidate Core Inflation Measures

Although the term "core inflation" has long meant an inflation series excluding food and energy price changes, alternative measures of core inflation have been proposed. This development likely reflects the lack of a widely accepted definition of core inflation. These alternative measures of core inflation are derived using one of two approaches. Borrowing the terminology of Mankikar and Paisley (2002), we refer to these methods as the "statistical approach" and the "model-based approach."

The statistical approach derives measures of core inflation by performing a predetermined operation on an aggregate price index. The operation may involve excluding certain items from the price index, re-weighting the components of the price index, or smoothing time-series movements in the price index. Alternatively, the model-based approach typically derives measures of core inflation by imposing restrictions from economic theory within the context of a multivariate econometric analysis. This approach leads to estimates of core inflation that may be associated with dynamic factor models or defined as a component of measured inflation possessing particular interactive effects with other variables.[8]

For this study, we restrict our attention to measures of core inflation associated with the statistical approach. We do this for several reasons. One is that the statistical approach yields core inflation measures that are more widely used by central banks and are more familiar to the public. Measures of this type often appear in central bank discussions of monetary policy or in the media. Another reason is that there is little consensus about the specification and identification schemes of model-based measures of core inflation. Last, there is a marked difference between the two approaches in terms of complexity. Model-based core inflation measures could remain problematic to policymakers and the public because the concepts underlying their design can be abstract and their construction computationally demanding. On the contrary, while there is a variety of core inflation measures associated with the statistical approach, each measure is relatively easy to understand and compute.

Within the statistical approach, the core inflation measures reflect very different characterizations of transitory price movements. Some of these measures associate the bulk of transitory price fluctuations with specific components, thereby prompting their exclusion from an aggregate price index. We consider two examples of this type of core inflation measure. One is based on the conventional practice of excluding food and energy price changes from movements in an aggregate series. The other has been proposed by Clark (2001), who argues for a core measure of inflation that removes only energy price changes from movements in an aggregate series. His motivation is that food prices, at least at the consumer level, likely react to many of the same forces that influence other retail prices, whereas energy price changes are dominated by transitory commodity price shifts.

As an alternative to core inflation measures that remove some prespecified item(s) in every period, Bryan and Cecchetti (1994) advance a measure that involves re-weighting all the components in the price index. Specifically, their proposed core inflation measure is the weighted median price change in a period, which is defined as the price change in the period for that product such that half the expenditure is for items whose prices are rising just as, or more, rapidly, and half is for items whose price changes are rising just as, or more, slowly. The weighted median is related to the "trimmed mean" concept of core inflation (Dolmas 2005).[9] Bryan and Cecchetti's argument for focusing on measures constructed along these lines is that

Although the term "core inflation" has long meant an inflation series excluding food and energy price changes, alternative measures of core inflation have been proposed. This development likely reflects the lack of a widely accepted definition of core inflation.

[8]As examples, Velde (2006) defines core inflation as the (unobserved) component common to a large number of individual price series, while Quah and Vahey (1995) define core inflation as the component of measured inflation that is uncorrelated with output at medium- to long-run horizons.

[9]The trimmed mean is the average price change computed when omitting a specified percentage of the highest and lowest price changes of products (weighted by their expenditure share) for a period. While our analysis does not include the trimmed mean measure among the candidate series, we believe that some caution needs to be exercised in evaluating this measure. Specifically, researchers typically use full-sample estimation techniques to determine how much the distribution of price changes should be trimmed. However, the use of a criterion function to optimally select the amount of trimming could favorably bias the performance of this measure within a particular period of interest. In our view, any evaluation of a trimmed mean measure should be undertaken using recursive estimation so that the trimmed mean is constructed sequentially. This method would circumvent any difficulties that arise from allowing the future history of the data to impact the construct of the series during an earlier time period.

the tails of the price distribution are mainly associated with temporary price level effects; thus, systematically eliminating their influence should yield a more robust measure of the persistent component of inflation.

In contrast to the weighted median that smoothes the cross-section of price changes, Cogley (2002) develops a core measure of inflation that down-weights past changes in the price index. His proposed core inflation measure involves the exponential smoothing of current and past aggregate price changes. The motivation for this measure is the idea that the government and private sector use adaptive methods to learn about a world in which there are occasional regime shifts in mean inflation.

For the analysis, we examine the following four candidate core inflation measures noted above:[10]

- the aggregate inflation series excluding food and energy,
- the aggregate inflation series excluding energy proposed by Clark (2001),
- the weighted median measure of the aggregate inflation series proposed by Bryan and Cecchetti (1994),[11] and
- the exponentially smoothed version of the aggregate inflation series proposed by Cogley (2002).

Cogley's formulation is given as:

$$\tilde{\pi}_t = g_0 \sum_{j=0}^{\infty} (1-g_0)^j \pi_{t-j}, \quad 0 < g_0 < 1, \tag{1}$$

where π denotes the relevant aggregate inflation measure. Equation 1 defines the core measure as a one-sided geometric distributed lag of current and past inflation. We follow Cogley and set the gain parameter $g_0 = 0.125$.

3.2 Performance Criteria

Previously, we argued that core inflation should be viewed as an intermediate target for an aggregate inflation goal. Using this proposition as a guide, we evaluate the candidate core measures of inflation based on criteria comparable to those discussed in Wynne (1999):[12]

1. *Transparency of construction*. It may be helpful to build a core price measure in a straightforward, relatively easy fashion. This criterion facilitates the communication of the concept in the policy dialogue.
2. *Similarity of means*. A core measure should have a mean comparable to the goal inflation series over a long period of time.
3. *Tracking the trend rate of inflation*. A core measure should display a close coherence to the underlying trend in the goal inflation series.
4. *Explanatory content*. A core measure should explain past movements in the goal inflation series as well as provide information about potential future developments.

It is important to note, however, that in the literature there has been little uniformity in the criteria used to evaluate core measures of inflation. For example, Cogley (2002) focuses on the within-sample regression fit of core inflation measures (part of criterion 4). Bryan and Cecchetti (1994) examine the marginal within-sample predictive content of core inflation measures as well as their out-of-sample forecast performance (criterion 4). Clark (2001) judges core inflation measures based on their complexity, similarity of means, ability to track a measure of the trend rate of inflation, and within-sample predictive content (criteria 1, 2, 3, and part of criterion 4). Consequently, our set of criteria listed above is not only consistent with the attributes considered in other studies, but also broader in scope.

It is important to note . . . that in the literature there has been little uniformity in the criteria used to evaluate core measures of inflation.

Given the lack of a common set of performance criteria for core inflation measures in the literature, a similar issue arises concerning the choice of the goal inflation series. Bryan and Cecchetti (1994), Clark (2001), Cogley (2002), and Blinder and Reis (2005) examine the (standard published) CPI, whereas Dolmas (2005) and Smith (2006) examine the PCE deflator, and Smith (2004) and Khettry and Mester (2006) examine both

[10] In addition to these four series, Rich and Steindel (2005) examine exponentially smoothed versions of the ex food and energy, ex energy, and weighted median series as candidate core inflation measures.

[11] The Federal Reserve Bank of Cleveland issues monthly estimates of the change in the weighted median CPI. The Bank has recently announced changes in the procedures used to construct this measure (Bryan and Meyer 2007) based on the work of Brischetto and Richards (2006). Our computations are based on the older procedure.

[12] Silver (2006) also discusses a wide range of comparable criteria for judging the relative merits of proposed core inflation measures. Wynne (1999), like Bryan and Cecchetti (1994), notes that at times the rationale for the construction of a core price index has been to identify the common component of price changes attributable to monetary policy. If such is the purpose of a core price index, however, then it is not altogether clear why one would confine the measure to elements of household price indexes. The difficulty is that monetary policy affects the demand for all types of products in complex ways. These demand effects are not necessarily similar for household and nonhousehold prices, nor is there any strong reason to assume that the distribution of monetary policy effects between household and other prices will be stable over time.

the CPI and the PCE deflator. One can also extend the list of differences among studies to include sample period, model specification, forecasting horizon, and data frequency. Thus, it seems unlikely under the circumstances that any type of consensus about core inflation measures would emerge, and the divergence of conclusions in the literature bears this out.

In light of the previous discussion, our study attempts to correct for the lack of standardization in the evaluation of core inflation measures. When it is feasible to encompass the features of other studies, such as in the selection of performance criterion and choice of goal inflation series, we do so. When it is not feasible (most notably, across sample period, model specification, forecast horizon, and data frequency), we rely on our judgment to ensure that these dimensions of the evaluation process are reasonable and similar to those adopted in other studies. Although we recognize that there are limits to the generality of our framework, we nevertheless believe that it offers an improved basis for judging the capabilities of core inflation measures, and that it may help to clarify the observed differences across previous studies.

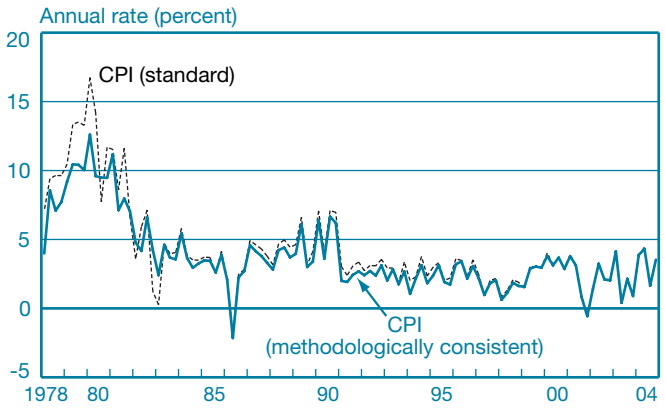
There is one additional point that merits attention, given the conflicting evidence reported in previous studies. Our discussion has emphasized how the design of core inflation measures is influenced by views about the nature of transitory price movements. It is important to note that these views reflect not only an explicit statement about the sources of transitory price movements, but also an implicit assumption concerning the invariance of these sources. If the pattern of transitory price movements were to change over time, then the reliability of core inflation measures would likely be affected. Moreover, if changes in the pattern of transitory price movements were to coincide with different sample periods used for estimation or forecasting purposes, then one would expect to observe variation in the relative performance of the core inflation measures across the sample periods. Keeping these considerations in mind, we now turn to the empirical framework.

4. Empirical Framework

For the analysis, we restrict our attention to aggregate inflation measures that would likely be of interest to policymakers and the public. We select two measures: quarterly growth in the PCE index and quarterly growth in the methodologically consistent CPI. Because the PCE index has gained considerable prominence in U.S. monetary policymaking at the expense of the CPI in recent years, we find that it is instructive to analyze the PCE in parallel with the CPI.[13]

The PCE index is produced while constructing the National Income and Product Accounts data, with the quarterly index

Chart 1
CPI Inflation: 1978-2004



Source: U.S. Department of Labor, Bureau of Labor Statistics.

excluding food and energy starting in 1959.[14] The methodologically consistent CPI is a less familiar price index. It is basically a reconstruction of the CPI designed to, as closely as possible, utilize current procedures to compute the prices of individual products. The major, but by no means sole, difference from the standard CPI is the extension of the current "rental equivalence" method of computing homeowners' costs to data prior to 1984. The key advantage of using such a series is that it controls for any impact on the statistical results that may arise from changes in the methods used to construct the CPI. A major disadvantage is that the series starts in 1978, limiting the time period available for the analysis.[15]

Not surprisingly, the aggregate methodologically consistent CPI series has differed from the standard published CPI series (Chart 1). The divergence is most notable, and quite striking, prior to 1984, when the measurement of homeownership costs in the standard CPI was changed to the owners' equivalent rent concept used for the entire history of the methodologically consistent series.

We now provide additional details on variable construction, metrics, model specifications, and testing procedures used in our analysis. The criteria used to evaluate the core inflation measures differ in terms of complexity. Whereas comparing the average rates of aggregate inflation and core inflation is relatively straightforward, the same can not be said for tracking the trend rate of inflation and either explaining or forecasting movements in aggregate inflation.

[13]For example, Federal Open Market Committee participants report their projections of inflation using the PCE index and the PCE index excluding food and energy, rather than the CPI excluding food and energy.

[14]We use the vintage of data available before the 2005 annual midyear revision.

[15]The methodologically consistent CPI is available monthly from the Bureau of Labor Statistics under the name "CPI-URX," and is often referred to as the "research series." We use the more cumbersome title to emphasize its advantages in statistical analysis. Stewart and Reed (1999) describe the index.

4.1 Tracking Trend Inflation

There are two ways to evaluate how well a core inflation measure tracks trend inflation. One way is based on the idea that a core inflation measure should neither understate nor overstate the long-run rate of growth of the goal price index. We can assess this feature by comparing the long-run means of a core inflation measure and aggregate inflation. Another way is based on the idea that a core inflation measure should match the movements in the trend rate of inflation over time. This assessment, however, requires additional assumptions about the way to estimate trend inflation and the metric to gauge the deviation between the series.

To construct the measure of trend inflation, we apply the Baxter-King (1999) band-pass filter to the data.[16] The band-pass filter returns a component that eliminates all periods less than eight years (thirty-two quarters).[17] To gauge the accuracy with which core inflation tracks trend inflation, we follow Clark (2001) and use a measure of volatility for this assessment. Specifically, we compute the root mean squared error (RMSE) of the difference between trend inflation and core inflation:

$$(2) \quad RMSE(\pi^{TREND} - \pi^{CORE}) = \sqrt{\sum_{t=1}^{T} (\pi_t^{TREND} - \pi_t^{CORE})/T},$$

where π_t^{TREND} is an estimate of the trend of inflation at time t and π_t^{CORE} is a particular measure of core inflation at time t.

4.2 Model Specification and Testing Procedures

Our final set of criteria addresses the ability of the candidate core inflation measures to account for movements in aggregate inflation both within sample and out of sample. The following specification serves as the benchmark model for this part of the analysis:

$$(3) \quad \pi_{t+h} - \pi_t = \alpha_h + \beta_h(\pi_t - \pi_t^{CORE}) + \varepsilon_{t+h},$$

where $\pi_{t+h} = (400) * ln(P_{t+h} / P_{t+h-1})$ is quarterly inflation h-quarters-ahead reported at an annual rate, $\pi_t = (400) * ln(P_t / P_{t-1})$ is current quarterly inflation reported at an annual rate, π_t^{CORE} denotes one of the indicators of current quarterly core inflation measured at an annual rate, and ε_{t+h} is a mean-zero random disturbance term.

The regression model in equation 3 has been used in studies such as Clark (2001), Hogan, Johnson, and Laflèche (2001), Cutler (2001), and Cogley (2002).[18] One attractive feature of the model is its easy interpretation. In particular, the model relates the change in inflation over the next h quarters to the contemporaneous gap between actual inflation and core inflation. That is, the current "core deviation" (transitory movement in inflation) is used to predict how much aggregate inflation will change over the next h quarters. The specification of the model accords with the intuition that if a core measure is identifying current price changes that are expected to dissipate, then the core deviation by definition should be providing a measure of an anticipated reversal in inflation.

Although the formulation of the model in equation 3 is admittedly simple, Clark (2001) and others argue that it is consistent with the beliefs of some policymakers and commentators who take movements in core inflation, by themselves, as signals of future changes in inflation. Moreover, the specification of the dependent and independent variables in terms of differences in inflation rates effectively ensures that the two variables are stationary, thereby circumventing any complications arising from the presence of unit roots.[19]

Our final set of criteria addresses the ability of the candidate core inflation measures to account for movements in aggregate inflation both within sample and out of sample.

Another attractive feature of the model is that we can draw upon the construct of a successful measure of core inflation to obtain restrictions on the parameters in equation 3. In particular, if one adopts the Bryan-Cecchetti (1994) definition of core inflation as "the component of price changes that is expected to persist over medium-run horizons of several years," then this relationship would imply:

$$(4) \quad \pi_t^{CORE} = E[\pi_{t+h} | I_t],$$

where E denotes the expectations operator and I_t is an

[16] Dolmas (2005) uses the Baxter-King band-pass filter to construct an estimate of trend inflation, while Bryan and Cecchetti (1994) and Clark (2001) use a centered moving average. Recently, Comin and Gertler (2006) have advocated the use of the Baxter-King filter against alternatives to examine movements in aggregate activity.

[17] While the band-pass filter is attractive for isolating components with particular periodicities, the estimate of a component at a point in time is based upon both past and future values of a series. Consequently, the band-pass filter is not designed to detect changes in trend inflation in real time and therefore would be of little value to a policymaker.

[18] Smith (2004, 2006) estimates models that are broadly comparable to equation 3, but her specifications include lagged values of actual inflation and core inflation as additional regressors.

[19] During the sample periods considered in this study, U.S. price inflation displays a very high degree of persistence. In particular, it is standard in the literature to model the series as a unit root process.

information set that includes information on price changes through time period t.[20] From equation 3, the Bryan-Cecchetti definition will hold under the joint restriction $\alpha_h = 0$ and $\beta_h = -1$. The value of β_h is of particular interest because it indicates whether the core deviation is correctly measuring the magnitude of the transitory movement in inflation. Specifically, a value of β_h greater than (less than) unity in absolute value indicates that the measured core deviation understates (overstates) the subsequent changes in inflation, and thus understates (overstates) the magnitude of current transients.

For the within-sample analysis, we undertake the estimation of equation 3, using all available observations over a sample period and allowing the values of h to range from one to twelve quarters. Whenever $h>1$, there will be overlapping observations caused by the forecast horizon exceeding the sampling interval of the data. Consequently, we use the Newey-West (1987) covariance matrix estimator to account for autocorrelation (and possible conditional heteroskedasticity) of the regression residuals.

In the case of the out-of-sample analysis, we generate the forecasts through recursive estimation of equation 3. Specifically, we use data through quarter $t+h$ to estimate the model, relating the change in the inflation rate between quarter $t+h$ and quarter t to the core deviation in quarter t. Although data on the core deviation for quarter $t+1$ through quarter $t+h$ are not used for estimation, these observations are part of the current information set. Consequently, the estimated model can be iterated forward by h quarters to produce an h-quarters-ahead forecast of inflation $\hat{\pi}_{(t+h)+h}$. We then move the sample forward by one quarter and repeat the exercise. For each measure of core inflation and horizon, the pseudo out-of-sample forecasting procedure will generate a single series of forecast errors $(\pi - \hat{\pi})$.[21] We can then compute the RMSE of forecasts to compare the performance of the measures of core inflation.[22]

The discussion up to this point has considered the core measures in isolation. However, we can augment the regression model to include other variables that may contain additional predictive content for future movements in inflation. In particular, we can extend equation 3 such that:

$$\pi_{t+h} - \pi_t = \alpha_h + \beta_h(\pi_t - \pi_t^{CORE}) + \gamma_h X_t + \varepsilon_{t+h}, \quad (5)$$

where X_t is a macroeconomic variable that is taken as an indicator of economic slack.[23] Despite a large set of possible candidates, we restrict ourselves to a measure of the unemployment rate, capacity utilization, and the output gap.[24] To adhere to the principle of parsimony, we experiment only with the macroeconomic variables on an individual basis so that γ_h is restricted to be a scalar. However, we consider both the level and first difference of the macroeconomic variables to account for the possibility of rate-of-change effects.

Regarding our evaluation procedure, there are two points that merit special discussion. The first is that we do not propose equation 3 as a model of how actual inflation forecasts are, or should be, prepared. Rather, equation 3 affords us an additional metric to rank the core measures based on their relative accuracy to forecast inflation within the benchmark regression model. A core measure that is better than others on this metric may be considered useful to communicate inflation risks in a straightforward manner, even if a central bank uses other means to construct its internal inflation forecast.

As a second point, we recognize that there may be caveats associated with some of the statistical criteria used to evaluate the candidate core inflation series. For example, the explanatory content of a core measure may vary over time. As documented by Cecchetti, Chu, and Steindel (2000), there is evidence that relationships that appear to satisfactorily predict inflation in one year can often deteriorate in the next. This consideration gains further relevance when the actions of a central bank are taken into account. Specifically, if a core inflation measure was informative about the inflation outlook, then a monetary authority might incorporate this result into its policy formulation. If this changed policy response led to greater stability of inflation, then conventional correlation measures would show a weakening in the link between the goal inflation series and the core inflation measure. Thus, an observed deterioration in the relationship could, paradoxically, result from a core inflation measure remaining a useful indicator of (potential) future developments in inflation.

Concerning this latter point, we examine the issue of stability of the core inflation measures across two dimensions. One dimension focuses on the forecast performance of the core inflation measures over alternative subperiods. The other

[20]For the moment, we can think of I_t as a subset of a larger information set that includes all data available through time t. The information set I_t is merely intended to represent the data used to construct the core measures of inflation examined in this study. As such, it would also include expenditure weights on the various components needed to compute the weighted median measures.

[21]We use the term "pseudo" to acknowledge the fact that the analysis does not use real-time data sets.

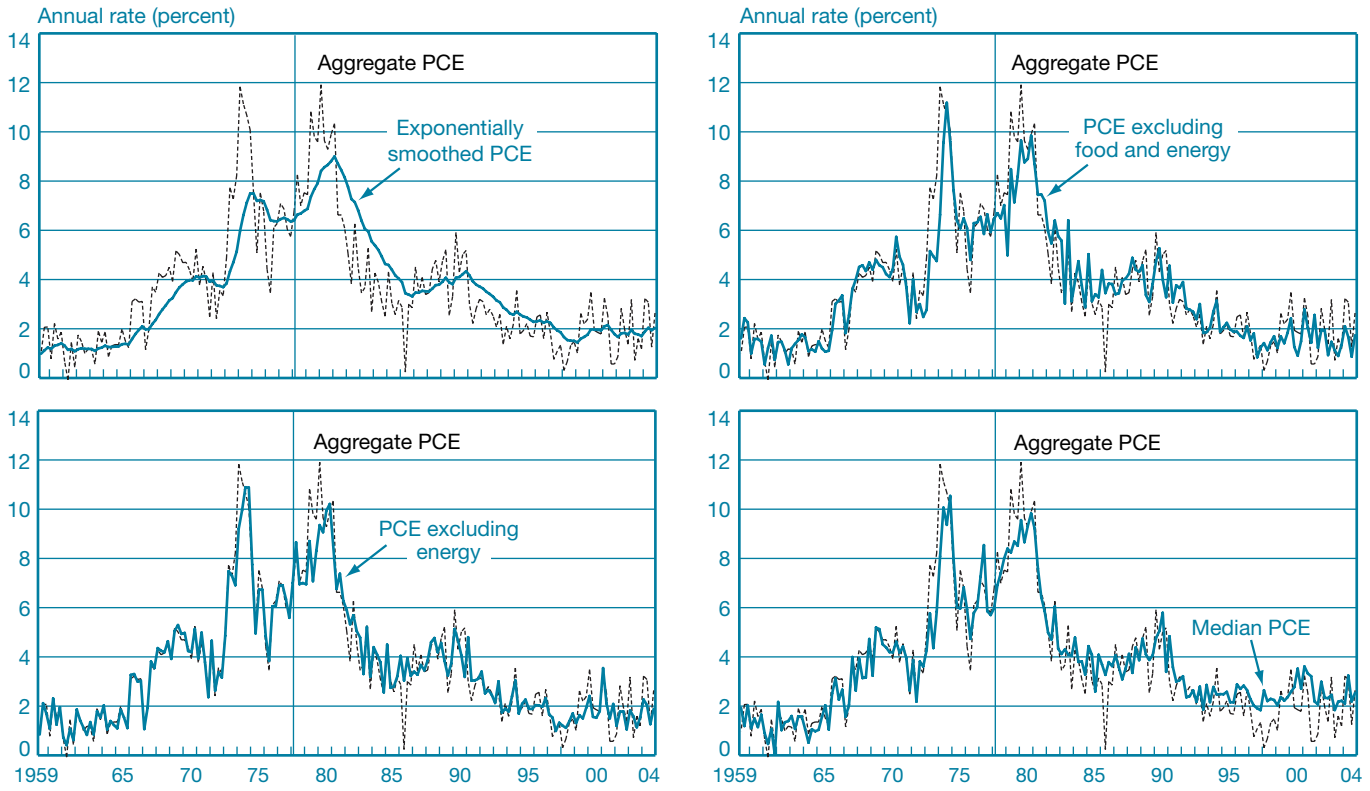
[22]In another study examining core measures of the methodologically consistent CPI, Marques, Neves, and Sarmento (2003) do not use RMSE to evaluate forecast performance. Instead, they argue that the deviation between aggregate inflation and a measure of core inflation should be correlated with future movements in aggregate inflation but uncorrelated with future movements in the core measure itself.

[23]We could have augmented X_t to include other types of variables such as financial indicators, oil prices, and various types of monetary aggregates. However, we selected measures of (excess) demand pressure in the economy based on the previous results of Cogley (2002) to keep the analysis manageable.

[24]The unemployment rate is for prime-age males (ages twenty-five to fifty-four) to control for demographic changes. Following Cogley (2002), we construct the output gap measure as $100 * (y_t - \tilde{y}_t)$, where y_t is the log of real GDP and $\tilde{y}_t$ is an estimate of the trend from applying the exponential smoother in equation 1. The measure of capacity utilization is for the manufacturing sector.

Chart 2

PCE Deflator Inflation and Core Measures: 1959-2004



Sources: U.S. Department of Commerce, Bureau of Economic Analysis; authors' calculations.

Note: The vertical bar at 1978:1 indicates the overlap of the various PCE and CPI series during the truncated sample period.

applies structural break tests to investigate the stability of the regression coefficients in equation 3.

5. Empirical Results

The data on the candidate core measures for PCE inflation start in 1959:2. For the methodologically consistent CPI, the observations on aggregate inflation and the candidate core measures start in 1978:1.[25] To provide a basis of comparison and to check the robustness of the results, we also undertake analysis of PCE inflation starting in 1978:1. Because the observations on the weighted median CPI series and the CPI ex energy series end in 2004:4, all of the analyses will end there to maintain consistency.

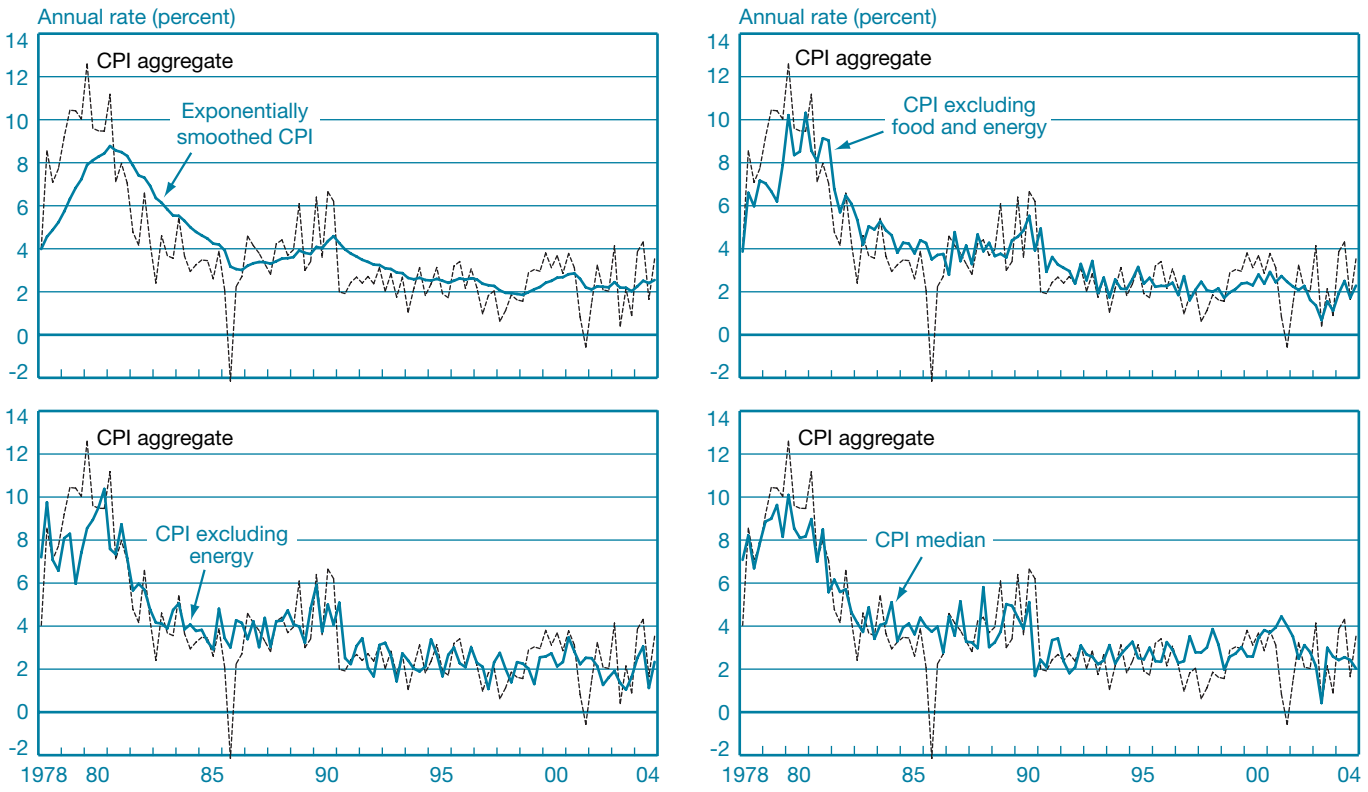
Chart 2 plots the growth of the PCE index against the candidate core measures over the period 1959-2004, while Chart 3 plots the methodologically consistent CPI over the period 1978-2004. Chart 2 contains a vertical bar at 1978:1 to illustrate the overlap of the various PCE series during the truncated sample period. Although the truncated sample period does not affect the values of actual PCE inflation or the measures excluding energy, excluding food and energy, and the weighted median, it slightly alters the initial values for exponentially smoothed series.[26]

[25]Quarterly values of the price indexes are averages of the relevant monthly figures. Quarterly growth rates are constructed from these averages. The exception is the growth rate of the CPI ex energy series for 1978:1, which is calculated as the growth of the index from December 1977 to March 1978.

[26]Our intention is to treat the PCE index and the CPI index in an identical manner over the truncated sample period. While the change in sample period does not affect the measures of economic slack such as the unemployment rate and capacity utilization, this is not true for our measure of the output gap. As indicated in footnote 24, the exponential smoother is also used to derive the output gap.

Chart 3

CPI Inflation and Core Measures: 1978-2004



Sources: U.S. Department of Labor, Bureau of Labor Statistics; authors' calculations.

5.1 Transparency of Design, Similarity of Means, and Coherence to Trend Inflation

All of our candidate core measures appear to be "transparent" in that they are related to the corresponding measures of aggregate inflation in a relatively straightforward, reproducible fashion.[27] Table 1 addresses the next two criteria for a core measure—the ability to match the mean of aggregate inflation and to track the trend rate of inflation. The table provides information on the PCE index for the post-1959 period and on the PCE index and the methodologically consistent CPI for the post-1978 period.

[27] Admittedly, the methodologically consistent CPI has not received a great deal of prominence, and the construction of the weighted median and exponentially smoothed measures involves a bit of effort. Because the current expenditure weights of PCE components may be read right off of the corresponding current-dollar consumption data, constructing the weighted median PCE series is a straightforward process.

The first criterion is similarity of means. As shown in Table 1, all the candidate core inflation measures have means quite near to that of aggregate PCE inflation during the full sample period. The situation is somewhat different during the post-1978 period. For both the weighted median series and the exponentially smoothed PCE, the means of the core inflation measures are somewhat higher than those of the respective aggregate inflation series.

Formal statistical tests revealed that the observed differences between long-run growth rates in aggregate prices and the core measures are statistically significant for the weighted median PCE and exponentially smoothed PCE during the 1978-2004 period. It is possible that the general process of disinflation that characterized much of the last quarter-century partly accounts for this finding. For example, the exponentially smoothed series are weighted sums of current and past inflation rates. Consequently, if inflation is generally trending downward, then there should be a tendency for the exponentially

TABLE 1

Average Inflation Rates and Volatilities around Trend Percent

	1959-2004	1961-2002
Inflation Measure	Mean[a]	RMSE($\pi^{TREND} - \pi^{CORE}$)[b]
PCE	3.69	—
PCE ex food and energy	3.60	1.03*
PCE ex energy	3.65	1.02*
Median PCE	3.77	1.00*
ES PCE	3.65	0.78
	1978-2004	1980-2002
Inflation Measure	Mean[a]	RMSE($\pi^{TREND} - \pi^{CORE}$)[b]
PCE	3.61	—
PCE ex food and energy	3.58	0.69
PCE ex energy	3.57	0.67
Median PCE	3.89**	0.79*
ES PCE	3.91**	0.73
	1978-2004	1980-2002
Inflation Measure	Mean[a]	RMSE($\pi^{TREND} - \pi^{CORE}$)[b]
CPI	3.80	—
CPI ex food and energy	3.86	0.73
CPI ex energy	3.80	0.76
Median CPI	4.03	0.89*
ES CPI	3.90	0.68

Source: Authors' calculations, based on data from the U.S. Department of Commerce, Bureau of Economic Analysis, and the U.S. Department of Labor, Bureau of Labor Statistics.

Notes: Test statistics were constructed using the Newey-West (1987) covariance matrix estimator. ES denotes an exponentially smoothed series.

[a] H_0: $\bar{\pi}^{AGGREGATE} = \bar{\pi}^{CORE}$.

[b] The Diebold-Mariano (1995) test statistic considers the null hypothesis of equal root mean squared error (RMSE) against the alternative hypothesis that the RMSE of a relevant benchmark series is lower.

*Statistically significant at the 5 percent level.

**Statistically significant at the 1 percent level.

smoothed measure of inflation to be a bit higher than current inflation. If inflation is declining, then there may also be a tendency for the outliers in the cross-sectional distribution of price change to be skewed toward negative or low readings, and thus contribute to the weighted median price change running a bit higher than the average price change.

The second criterion is the proximity of a candidate core measure to the current underlying trend in the goal inflation series. Chart 4 depicts the trend estimates, from which we drop two years of data from the beginning and end of the band-pass filtered series because they are relatively poorly estimated. Turning back to Table 1, we note that the RMSEs show that the exponentially smoothed measures tend to track the trend more closely for CPI inflation and PCE inflation during the longer sample, whereas the ex energy measure tracks the trend in PCE inflation more closely during the shorter sample.

Similar to the analysis comparing long-run growth rates, we can conduct formal tests to determine if the observed differences in RMSE are statistically significant. To address this issue, we construct the Diebold-Mariano (1995) test statistic, which allows us to consider the null hypothesis that two core inflation measures track the trend rate of inflation equally well against the alternative hypothesis that one core inflation measure tracks trend inflation more accurately than the other. For the tests, we select the core inflation measure associated with the lowest RMSE as the benchmark series and then make comparisons with the other core measures on an individual basis. We employ an error criterion for the test based on the difference in squared prediction errors across the two core inflation measures:

$$d_{t+i} = (\hat{e}^2_{1,\,t+1} - \hat{e}^2_{1,\,t+1}), \tag{6}$$

where $\hat{e}^2_{j,\,t+i} = (\pi^{TREND}_{t+i} - \pi^{CORE}_{j,\,t+i})^2$ and d_{t+i} is the differential loss in period $t+i$ using the benchmark series ($j=1$) versus the alternative series ($j=2$). The Diebold-Mariano test essentially determines whether the mean differential loss $\bar{d}$ across a selected sample period is statistically different from 0.

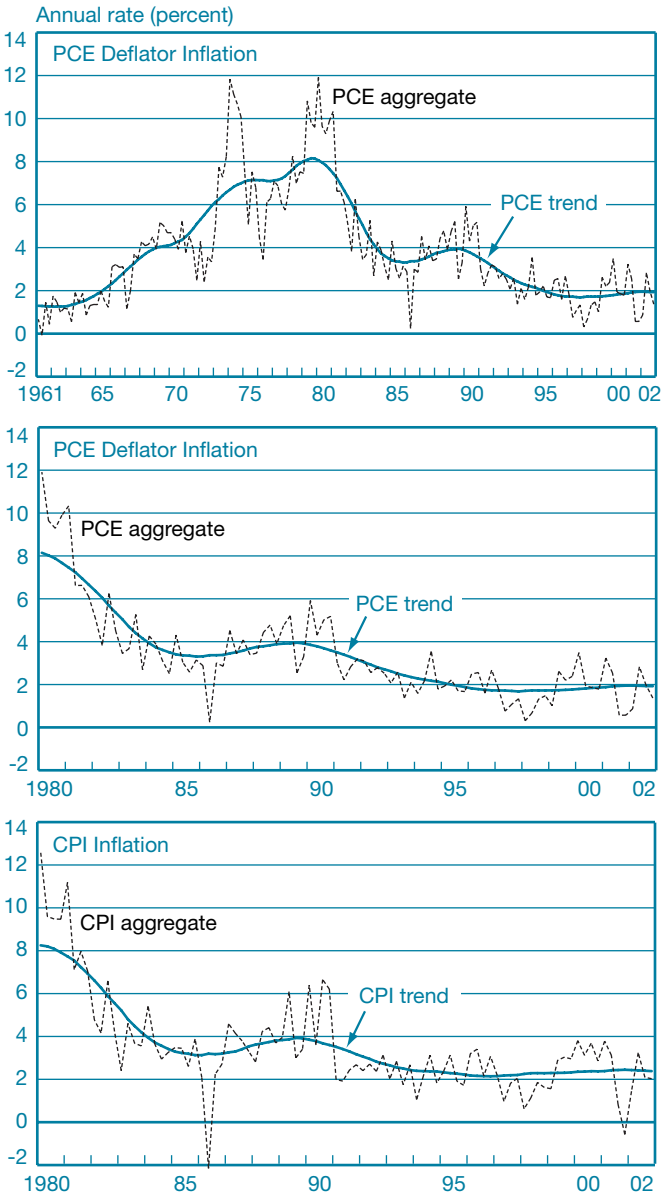
The Diebold-Mariano tests yield similar results during the shorter sample period. The ex energy measure performs significantly better than the weighted median in tracking trend PCE inflation, whereas the exponentially smoothed measure performs significantly better than the weighted median in tracking trend CPI inflation. For the longer sample, however, the exponentially smoothed measure performs significantly better than each alternative core measure in tracking trend PCE inflation. This latter result is driven principally by an episode from the early and mid-1970s in which there was a pronounced deviation between trend PCE inflation and the ex food and energy series, ex energy series, and weighted median series.[28]

All of our candidate core measures appear to be "transparent" in that they are related to the corresponding measures of aggregate inflation in a relatively straightforward, reproducible fashion.

[28] The reason for the deviation is that trend PCE inflation was much higher than these three core inflation series during this episode.

Chart 4

Aggregate Inflation and Estimated Trend Inflation Using Band-Pass Filter



Sources: U.S. Department of Commerce, Bureau of Economic Analysis; authors' calculations.

An initial read of the findings suggests that the performance of the core inflation measures is fairly similar on balance. While the exponentially smoothed PCE series is much better than all other core inflation measures at tracking trend inflation over the longer sample, its performance does not carry over to either CPI or PCE inflation during the shorter sample. Moreover, the exponentially smoothed PCE series actually overstates the average rate of PCE inflation during the shorter sample. Even though there is no superior core inflation measure that emerges from Table 1, one could argue that the weighted median series is distinguished by its relatively poor performance in tracking trend inflation. One explanation for this feature of the weighted median may be its re-weighting of the individual price series and thus the weights on their relative trends. Such re-weighting also occurs in the construction of the ex food and energy and the ex energy measures, but it is more extensive for the weighted median.

5.2 Within-Sample Evaluation

We now examine the results from estimating equation 3 across the various sample periods, forecast horizons, and core measures. Because of the large number of regressions and our desire to conserve space in the reporting of results, we devote the bulk of our discussion to the results using the (methodologically consistent) CPI as the inflation measure, and then summarize the findings for PCE inflation. In a further effort to save space, we generally do not report estimates of individual parameters and standard errors. More complete results may be found in Rich and Steindel (2005).

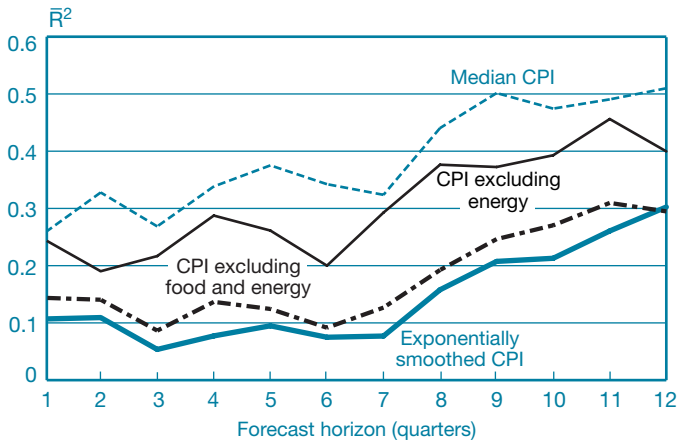
Chart 5 plots the goodness of fit, as measured by the $\bar{R}^2$ statistic, for the core inflation measures over horizons that range from one to twelve quarters. Some of the core measures appear to explain a significant amount of CPI growth over these horizons. This is particularly true for the weighted median, which may not fare well in terms of tracking trend inflation but is able to explain approximately 50 percent of the total variation in CPI inflation at horizons exceeding eight quarters (two years). Alternatively, it is interesting to note that the conventional ex food and energy measure and the exponentially smoothed measure tend to have the lowest explanatory content for CPI inflation.

In addition, there is a marked increase in the explanatory content of the core measures as the forecast horizon increases beyond six quarters. This result corroborates the previous findings of Hogan, Johnson, and Laflèche (2001) and Cogley (2002) and is consistent with the intended design of a core measure to identify transients in the data. Not surprisingly, one would expect a greater effectiveness at filtering transitory price changes to translate directly into an improved ability to explain subsequent reversals in inflation.

Table 2 presents tests for unbiasedness ($\alpha_h = 0$ and $\beta_h = -1$). These tests were conducted only at the twelve-quarter horizon to allow sufficient time for the identified transients to dissipate. We strongly reject unbiasedness for the weighted median measure, with the test statistic for the ex energy measure

Chart 5

CPI Inflation: 1978-2004
Within-Sample Fit



Source: Authors' calculations, based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

Table 2

Unbiasedness Test

Core Inflation Measure	α_{12}^a	β_{12}^b	$H_0: \alpha_{12} = 0$ and $\beta_{12} = -1$
CPI ex food and energy	-0.973	-1.024	3.817
	(0.507)	(0.166)	p-value = 0.148
CPI ex energy	-0.906*	-1.231	5.658
	(0.440)	(0.153)	p-value = 0.059
Median CPI	-1.161**	-1.332	8.678
	(0.411)	(0.263)	p-value = 0.013
ES CPI	-0.936	-0.885	3.139
	(0.529)	(0.224)	p-value = 0.208

Source: Authors' calculations, based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

Notes: Standard errors, reported in parentheses, were calculated using the Newey-West (1987) covariance matrix estimator. ES denotes an exponentially smoothed series.

[a] $H_0: \alpha_{12} = 0$.
[b] $H_0: \beta_{12} = -1$.

*Statistically significant at the 5 percent level.
**Statistically significant at the 1 percent level.

just slightly below the critical value at the 5 percent significance level.[29] It is interesting to note that the source of bias does not stem from the slope coefficient, as all of the core measures satisfy the individual restriction $\beta_{12} = -1$ at conventional significance levels. Rather, the bias is related to the intercept displaying a large and negative value. This result is not particularly surprising, given the behavior of inflation over the post-1980 period. Specifically, it would appear that the core measures are effectively removing the transient noise around trend inflation, although some measures are unable to account for the sustained decline in trend inflation over this sample period.

As discussed previously, there is no reason to confine our attention to measures of core inflation when explaining subsequent movements in aggregate inflation. We now consider the results from estimating equation 5, in which we combine the core measures with macroeconomic variables that are conventionally viewed as indicators of slack in the economy. Because of the even larger number of estimated regressions involved in this part of the analysis, we again elect to provide a summary of the main findings.

These results suggest four main conclusions:

First, the general features associated with the predictability of inflation carry over from the univariate analysis. That is, the $\bar{R}^2$ for the combinations of core inflation and macroeconomic variables tends to rise as the horizon increases.

Second, although the addition of macroeconomic variables can improve the predictability of aggregate inflation, their contribution can vary considerably across the core measures.

[29] The test statistic is distributed asymptotically as a chi-square random variable with two degrees of freedom.

The top and bottom panels of Chart 6 depict the (adjusted) goodness-of-fit measures for both the univariate regressions and bivariate regressions for the weighted median CPI and exponentially smoothed CPI, respectively. As shown in the top panel, the macroeconomic variables offer very little improvement in the fit of the regression over the various horizons.[30] In contrast, the bottom panel shows a marked improvement in the fit of the regression when we include capacity utilization, the output gap, or the unemployment rate as a second explanatory variable.

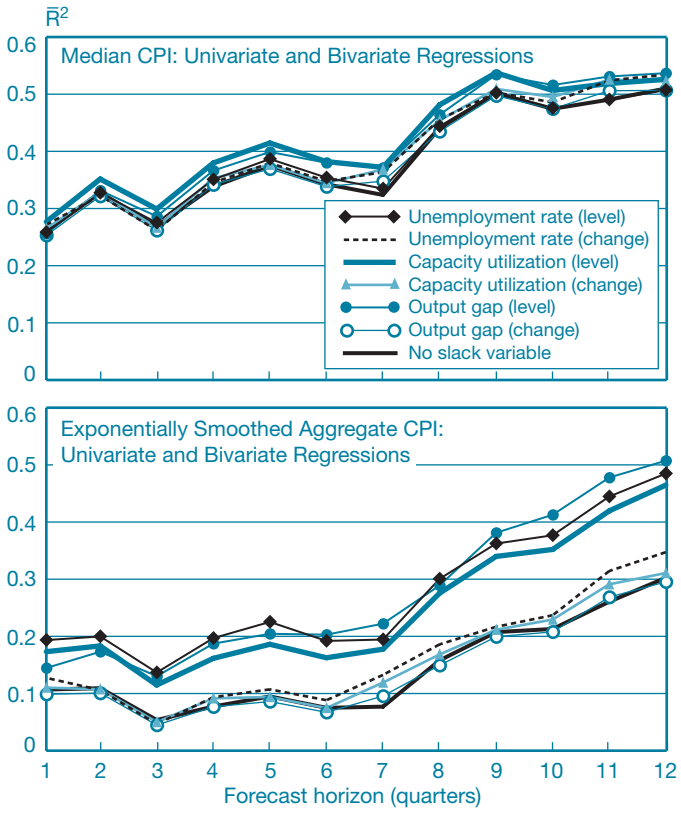
Third, the predictive power of the core measures, in combination with the macroeconomic variables, was almost always higher when the macroeconomic variables entered the regression in level form rather than as a first difference. We interpret this finding as indicating the absence of "rate-of-change" effects. The bottom panel of Chart 6 is representative of these findings. Within the candidate list of macroeconomic variables, the level of the unemployment rate typically resulted in the largest improvement in the fit of the regression equation.

Finally, the highest $\bar{R}^2$ for the CPI inflation regressions was associated with the ex energy and weighted median measures (along with the level of a macroeconomic variable). This latter finding contrasts sharply with that of Cogley (2002) and likely reflects our use of the methodologically consistent CPI series as well as a difference in sample periods.

[30] Admittedly, this statement may require some qualification due to the fairly impressive fit of the univariate regression on its own.

Chart 6

CPI Inflation: 1978-2004



Source: Authors' calculations, based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

5.3 Out-of-Sample Forecast Evaluation

For the out-of-sample analysis, we report the results only for horizons corresponding to four quarters, eight quarters, and twelve quarters. These horizons seem to be of particular interest to central banks, and again allow us to keep the discussion manageable.[31] We select two starting dates for our out-of-sample forecasts: 1990:1 and 1995:1. That is, the first four-quarter out-of-sample forecast corresponds to 1990:4.[32] Similar to the within-sample analysis, all out-of-sample forecast analyses end in 2004:4.

[31]Rich and Steindel (2005) also report results for a one-quarter forecast horizon.

[32]As we discussed, this forecast is based on model estimation for the four-quarter-ahead change in the inflation rate up through 1989:4 using data on the regressors for 1988:4 and earlier. The forecast is then constructed by iterating the model forward by four quarters and using the available observations on the regressors for 1989:4.

Table 3

Forecasting Performance of Alternative Measures of Core Inflation: RMSE of Univariate Forecasts

	Post-1990 Sample			Post-1995 Sample		
	RMSE $(\pi - \pi)$					
Core Inflation Measure	$h=4$	$h=8$	$h=12$	$h=4$	$h=8$	$h=12$
CPI ex food and energy	1.523	1.480	1.703**	1.483	1.591	1.885
CPI ex energy	1.432	1.555	1.825**	**1.382**	1.574	1.867
Median CPI	**1.383**	1.477	1.868	1.491	**1.467**	**1.658**
ES CPI	1.586	**1.455**	**1.506**	1.518	1.607	1.752

Source: Authors' calculations, based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

Notes: The Diebold-Mariano (1995) test statistic considers the null hypothesis of equal root mean squared error (RMSE) against the alternative hypothesis that the RMSE of a relevant benchmark series is lower. Test statistics were constructed using the Newey-West (1987) covariance matrix estimator. Figures in bold are the lowest RMSE at each horizon for the various out-of-sample forecast periods. ES denotes an exponentially smoothed series.

*Statistically significant at the 5 percent level.
**Statistically significant at the 1 percent level.

We restrict our discussion to the univariate forecasts based on the regression model in equation 3. Table 3 reports the RMSE statistics, in annual percentage terms, for the post-1990 and post-1995 periods. We highlight in bold the lowest RMSE at each horizon for the various out-of sample forecast periods.

Taken together, the evidence indicates that the forecast performance of the core measures, in both relative and absolute terms, can vary over the choice of sample period. It also varies across forecast horizons, although the differences are more noticeable and emerge more clearly over the medium run of several years. The exponentially smoothed measure delivers the lowest RMSE at the longer horizons in the post-1990 sample. In contrast, in the post-1995 sample, the weighted median measure does well, and the relative performance of the exponentially smoothed measure deteriorates noticeably. For example, the exponentially smoothed measure had the lowest RMSE for the entire post-1990 period at the eight-quarter (two-year) horizon, but the highest RMSE at the same horizon during the post-1995 period. In addition, the ex energy measure performed quite poorly at the longer forecast horizons. This finding also contrasts with Clark's (2001) reported evidence and, again, likely reflects differences in sample periods as well as in our use of the methodologically consistent CPI series.

As before, we can apply Diebold-Mariano tests to assess if the observed differences in RMSE in Table 3 are statistically significant. When we use conventional significance levels, the evidence indicates that there is almost no difference in forecast performance of the core inflation measures across the various horizons during either sample period. The one exception is the exponentially smoothed series during the post-1990 period, which is more accurate than the ex food and energy series and the ex energy series at the twelve-quarter horizon, although this result does not carry over to the weighted median series. This latter finding is somewhat interesting because of the higher RMSE associated with the weighted median, suggesting that it generally compares favorably with the exponentially smoothed series but is occasionally subject to large forecast errors.

Combining the evidence from the within-sample and out-of-sample analyses, we still conclude that there is no clearly dominant core CPI inflation measure. All measures failed to satisfy the test for unbiasedness. Whereas the weighted median possessed superior within-sample explanatory power for future movements in inflation that is also economically significant, it understated the degree of inflation reversal during the 1978-2004 period. In terms of the exponentially smoothed series, a slightly favorable forecast capability was offset by extremely low within-sample explanatory power. With regard to the ex energy series, and especially the ex food and energy series, these measures were not very noteworthy.

Table 4

Summary of Within-Sample and Out-of-Sample Analyses: Best-Performing Measures of Core Inflation

Within-Sample Fit and Range of $\bar{R}^2$	PCE: 1959-2004	PCE: 1978-2004
$h = 4$	Ex energy (0.014 - 0.081)	Median (0.043 - 0.173)
$h = 8$	Ex energy (0.102 - 0.185)	Median (0.142 - 0.345)
$h = 12$	ES (0.142 - 0.177)	Median (0.252 - 0.366)

Out-of-Sample RMSE $(\pi - \hat{\pi})$	Horizon	PCE: 1959-2004	PCE: 1978-2004
1980-2004	$h = 4$	Ex energy	—
	$h = 8$	Ex energy	—
	$h = 12$	Ex energy	—
1990-2004	$h = 4$	Ex energy	Ex energy
	$h = 8$	Ex food and energy	Median
	$h = 12$	Ex energy	Median
1995-2004	$h = 4$	Ex energy	Median
	$h = 8$	ES	Median
	$h = 12$	Ex food and energy	Median

Source: Authors' calculations, based on data from the U.S. Department of Commerce, Bureau of Economic Analysis.

Note: RMSE is root mean squared error; ES denotes an exponentially smoothed series.

5.4 Results for PCE Inflation

The benchmark model for the PCE index was estimated using both long (1959-2004) and short (1978-2004) sample periods. Table 4 summarizes the findings for the within-sample and out-of-sample analyses. We indicate which core inflation measure generates the highest $\bar{R}^2$ for the benchmark regression model at horizons of four, eight, and twelve quarters, and we also report the range of the within-sample fit across all the core inflation measures. These results are followed by a listing of the core inflation measure that provides the most accurate out-of-sample forecast at these horizons, with the forecasts starting in 1990 and 1995. We also include forecasts starting in 1980 for the PCE deflator over the longer sample period.

Over the long sample period, the core inflation measures showed a remarkably similar pattern in terms of their explanatory content for future inflation. In particular, the within-sample fit of equation 3 was quite low, with the $\bar{R}^2$ statistic peaking at around 0.20 at the eight-quarter horizon and then remaining fairly constant. With regard to the parameter restrictions $\alpha_h = 0$ and $\beta_h = -1$, we were unable to reject them either individually or jointly at the twelve-quarter horizon. In contrast to the results for the CPI, we found that all of the core PCE inflation measures possess the property of unbiasedness over the long sample period.

Over the short sample period, the within-sample fit of equation 3 tended to increase with the horizon and was broadly comparable, albeit a bit lower, to those for the CPI regressions. Similar to our findings for the CPI, the weighted median had the most explanatory content for overall inflation across the horizons, with the exponentially smoothed series and ex energy series displaying the least explanatory content. The inclusion of (the level of) macroeconomic variables generally improved the within-sample fit of equation 5 for each core measure, although the magnitude of the improvement was more modest compared with that shown over the long sample. In fact, the $\bar{R}^2$ of the bivariate regression for each measure of core PCE inflation now displays a greater similarity across the two

sample periods. As was the case for the CPI, the short sample period reveals some evidence of bias in the core inflation measures at the twelve-quarter horizon. We strongly reject unbiasedness for the weighted median, with the test statistics for the ex energy and exponentially smoothed measures just slightly below the critical value at the 5 percent significance level. Moreover, the estimated intercept for all four measures was large, negative, and statistically significantly different from zero at conventional significance levels.

Out-of-sample forecast results (including those starting in 1980 for the PCE regression estimated over the post-1959 period) again varied substantially. When the benchmark model was estimated with data starting in 1959, the ex energy measure typically provided the most accurate forecasts over the post-1980 and post-1990 periods. From a statistical standpoint, however, the relative forecast performance of the ex energy measure was most meaningful at the four-quarter horizon and then tended to dissipate as the forecast horizon increased. In the case of the forecasts for the post-1995 period, the core inflation measure associated with the lowest RMSE varied from the ex energy series (four-quarter horizon) to the exponentially smoothed series (eight-quarter horizon) to the ex food and energy series (twelve-quarter horizon). In most cases, the difference in RMSE across the core inflation measures was not statistically significant.

Conversely, we observed a much more consistent pattern when we examined the out-of-sample forecast results based on estimating the benchmark model with data starting in 1978. For the post-1990 period, the ex energy series produced the lowest RMSE at the four-quarter horizon. In all other cases, however, the weighted median measure produced the most accurate forecasts. Moreover, the forecast performance of the weighted median relative to the other core inflation measures was statistically meaningful at both the eight- and twelve-quarter horizons over the post-1995 period. It is also worth noting that, for each forecast horizon over the post-1990 and post-1995 periods, the ex food and energy series produced the highest RMSE in five out of six possible cases.

The findings for the PCE deflator as well as those presented earlier for the CPI offer little, if any, compelling evidence in support of a preferred measure of core inflation. In many instances, the performance of the candidate series looks roughly comparable. On those occasions in which one series displays superior performance, it sometimes does not carry over to changes in the measure of aggregate inflation or the periods used for estimation and forecasting purposes. When there is evidence indicating that a core inflation measure may be well suited for performing a particular task, the same measure often displays inferior performance in terms of other criteria.

5.5 A Closer Look at the Results: Instability in the Core Inflation Measures, or Something Else?

Even though our findings contrast with claims, either explicit or implicit, in support of a particular measure of core inflation, other studies have drawn conclusions similar to ours. Using Canadian price data, Hogan, Johnson, and Laflèche (2001) examine several proposed measures of core inflation. They find that the candidate series are quite similar in many respects, but there is also evidence that some measures fare better than others along different dimensions. Rather than selecting one measure to perform the role of core inflation, they recommend examining a limited number of measures and using the varied information to assess the inflation outlook. Mankikar and Paisley (2002), examining price data for the United Kingdom, also find that there is no single measure of core inflation that performs well across a set of performance criteria.

Given our inability to identify a preferred core inflation measure, a reasonable reaction might be to look for instability in the relationship between aggregate inflation and the core inflation measures. We alluded to this concern previously in the context of a monetary authority that changed its behavior by incorporating information from a core inflation measure to set future policy.[33] In addition, some commentators have claimed that a wide range of inflation forecasting models experienced a breakdown during the mid-1990s. Further, a comparison of the goodness-of-fit and RMSE results reported for PCE inflation across the longer and shorter samples would appear to hint at the possibility of parameter instability.

To explore this issue more formally, we apply the predictive tests developed by Ghysels, Guay, and Hall (1998) to investigate the stability of the benchmark model. This testing procedure is particularly attractive because it does not require the researcher to specify the break point a priori, but rather allows the data to determine the break point. Because the disparity in the performance of the core measures is more notable at longer horizons, we restrict our attention to horizons of $h = 8$ and 12 quarters.[34]

Details of the test results can be found in Rich and Steindel (2005). In summary, the parameters of equation 3 for PCE inflation over the longer sample period appear stable using

[33]Of course, the issue of stability of these types of measures would be relevant in any situation where the economy undergoes a significant structural change.

[34]We are grateful to Arturo Estrella and Tony Rodrigues for providing the computer program used for this testing procedure. The joint tests of parameter stability impose a common break date across the parameters. The testing procedure also requires that we exclude from consideration a fraction of the sample at each end, so that the set of possible break points lies within an interior range. Consequently, our testing procedure is based on a 5 percent trimming of the sample. Because of the overlapping nature of the data, we again employ the Newey-West (1987) covariance matrix estimator in the construction of the test statistics.

conventional significance levels. For the post-1978 period, the test statistics for PCE inflation also provide little evidence of parameter instability. For CPI inflation, the results speak to more concern about parameter instability, but the evidence is still not very compelling in terms of incidence and statistical significance. Overall, it does not appear that instability in the structure of the underlying estimating model contributes significantly to our results.

As an alternative to the idea of a breakdown in the relationship between aggregate inflation and the core measures, we offer a more intrinsic explanation for our inability to identify a preferred core inflation measure. Simply stated, it is our view that there is too much variability in the nature and sources of transitory price movements to be effectively captured through the design of any one of the core inflation measures. For example, although there may normally be large transitory elements in movements in energy prices or food prices or both, the sustained run-up over the last few years in oil (and corn) prices highlights episodes in which these movements have displayed a greater degree of persistence. Similarly, the validity of core inflation measures based on a truncation of the distribution of price changes relies on the premise that large price changes—either increases or decreases—in individual items are generally associated with temporary price level effects.[35] As Mankikar and Paisley (2002) note, however, theoretical arguments based on menu cost models and staggered-price-setting models argue that both the excluded tails and the included portion of the price-change distribution contain a mixture of transitory and permanent shocks. Likewise, it is entirely reasonable to argue that there have been regime shifts in the mean of U.S. inflation that would justify a down-weighting of past price changes along the lines of the exponentially smoothed measure. However, the usefulness of this particular core inflation measure would seem less compelling during sustained periods of time characterized by intra-regime rather than inter-regime movements in inflation.

Simply stated, it is our view that there is too much variability in the nature and sources of transitory price movements to be effectively captured through the design of any one of the core inflation measures.

While a more detailed characterization of transitory price movements is beyond the scope of this article, we suggest that the lack of a consistent pattern to the movements' behavior could also help account for the differential performance of the core inflation measures documented in other studies. If there is variation in the nature and sources of transitory price movements, then different choices of models, sample periods, and criteria would be expected to yield different results. Thus, we would argue that the diversity of findings simply reflects the confluence of the lack of a consistent pattern to temporary price movements and the lack of standardization in the evaluation process.

[35]The validity of this type of core inflation measure also depends on the ability to specify the threshold defining the magnitude of a large price change.

6. Conclusion

Viewing the stabilization of CPI or PCE inflation as plausible goals for U.S. monetary policy, we evaluate several proposed measures of "core" inflation. Other studies have addressed this issue, but there has been little commonality in their underlying approaches. Thus, a key feature of our analysis is to provide a more consistent basis on which to judge the performance of the candidate core inflation series. This consideration guided us in our selection of model specification, criteria, and sample periods.

Given our empirical framework and greater standardization in the evaluation process, one possible outcome from our study was that a single core inflation measure would emerge as dominant in its performance. This, however, was not the case. Rather, we documented considerable variation in the performance of the candidate series. Further, we noted the rather unremarkable performance of the conventional ex food and energy series. Consequently, the general practice of identifying "core inflation" with an ex food and energy series instead of an alternative series does not seem to be justified based on our analysis.

Although the results of this study do not rule out the potential usefulness of core inflation measures, there appear to be difficulties associated with how best to employ the current set of candidates. One possibility is to weight various criteria and then select the core inflation measure that yields the best performance. However, this approach would be influenced by the highly subjective process of ranking the importance of the criteria. Another possibility is to acknowledge that different core inflation measures seem better suited to performing different tasks, and then adopt the appropriate core inflation measure as the guide for a particular stated purpose. However, this approach would introduce the inconvenience of keeping track of a variety of core inflation measures; moreover, in the policy area, it could require that a central bank provide the public with a clear understanding of each series. Finally, a central bank could consider the adoption of a model-based

measure of core inflation. However, this approach would then face the previously cited difficulties of choosing the particular definition of core inflation and communicating it to the public. Taken together, these considerations and our results present challenging avenues for future research.

References

Aoki, K. 2001. "Optimal Monetary Policy Response to Relative Price Changes." Journal of Monetary Economics 48, no. 1 (August): 55-80.

Baxter, M., and R. G. King. 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." Review of Economics and Statistics 81, no. 4 (November): 575-93.

Benigno, P. 2004. "Optimal Monetary Policy in a Currency Area." Journal of International Economics 63, no. 2 (July): 293-320.

Blinder, A. S., and R. Reis. 2005. "Understanding the Greenspan Standard." In The Greenspan Era: Lessons for the Future. Symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 25-27.

Brischetto, A., and A. J. Richards. 2006. "The Performance of Trimmed Mean Measures of Underlying Inflation." Reserve Bank of Australia Research Discussion Paper no. 2006-10, December.

Bryan, M. F., and S. G. Cecchetti. 1994. "Measuring Core Inflation." In N. G. Mankiw, ed., Monetary Policy: NBER Studies in Business Cycles, vol. 29, 195-215. Chicago and London: University of Chicago Press.

Bryan, M. F., and B. H. Meyer. 2007. "Methodological Adjustments to the Median and 16 Percent Trimmed-Mean CPI Estimates." Federal Reserve Bank of Cleveland. Available at <http://www.clevelandfed.org/research/inflation/US-Inflation/revmcpi.pdf>.

Cecchetti, S. G., R. S. Chu, and C. Steindel. 2000. "The Unreliability of Inflation Indicators." Federal Reserve Bank of New York Current Issues in Economics and Finance 6, no. 4 (April).

Clark, T. E. 2001. "Comparing Measures of Core Inflation." Federal Reserve Bank of Kansas City Economic Review 86, no. 2 (second quarter): 5-31.

Cogley, T. 2002. "A Simple Adaptive Measure of Core Inflation." Journal of Money, Credit, and Banking 34, no. 1 (February): 94-113.

Comin, D., and M. Gertler. 2006. "Medium-Term Business Cycles." American Economic Review 96, no. 3 (June): 523-51.

Cutler, J. 2001. "Core Inflation in the U.K." Bank of England External MPC Unit Discussion Paper no. 3, March.

Diebold, F. X., and R. Mariano. 1995. "Comparing Predictive Accuracy." Journal of Business and Economic Statistics 13, no. 3 (July): 253-63.

Dolmas, J. 2005. "Trimmed Mean PCE Inflation." Federal Reserve Bank of Dallas Working Paper no. 0506, July.

Dotsey, M., and T. Stark. 2005. "The Relationship between Capacity Utilization and Inflation." Federal Reserve Bank of Philadelphia Business Review, second quarter: 8-17.

Eckstein, O. 1981. Core Inflation. Englewood Cliffs, N.J.: Prentice-Hall.

Ghysels, E., A. Guay, and A. Hall. 1998. "Predictive Tests for Structural Change with Unknown Breakpoint." Journal of Econometrics 82, no. 2 (February): 209-33.

Goodfriend, M. S., and R. G. King. 1997. "The New Neoclassical Synthesis and the Role of Monetary Policy." In B. S. Bernanke and J. Rotemberg, eds., NBER Macroeconomics Annual 1997, 231-83. Cambridge: MIT Press.

Gordon, R. J. 1975a. "Alternative Responses of Policy to External Supply Shocks." Brookings Papers on Economic Activity, 1975, no. 1: 183-206.

———. 1975b. "The Impact of Aggregate Demand on Prices." Brookings Papers on Economic Activity, 1975, no. 3: 613-70.

Hogan, S., M. Johnson, and T. Laflèche. 2001. "Core Inflation." Bank of Canada Technical Report no. 89, January.

Khettry, N. K., and L. J. Mester. 2006. "Core Inflation as a Predictor of Total Inflation." Federal Reserve Bank of Philadelphia Research Rap, Special Report, April 26.

Mankikar, A., and J. Paisley. 2002. "What Do Measures of Core Inflation Really Tell Us?" Bank of England Quarterly Bulletin, winter: 373-83.

Marques, C. R., P. D. Neves, and L. M. Sarmento. 2003. "Evaluating Core Inflation Indicators." Economic Modelling 20, no. 4 (July): 765-75.

References (Continued)

Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." Econometrica 55, no. 3 (May): 703-8.

Quah, D., and S. V. Vahey. 1995. "Measuring Core Inflation." Economic Journal 105, no. 432 (September): 1130-44.

Rich, R., and C. Steindel. 2005. "A Review of Core Inflation and an Evaluation of Its Measures." Federal Reserve Bank of New York Staff Reports, no. 236, December.

Silver, M. 2006. "Core Inflation Measures and Statistical Issues in Choosing among Them." IMF Working Paper no. WP/06/97, April.

Smith, J. K. 2004. "Weighted Median Inflation: Is This Core Inflation?" Journal of Money, Credit, and Banking 36, no. 2 (April): 253-63.

———. 2006. "PCE Inflation and Core Inflation." Unpublished paper, Lafayette College, July 6.

Stewart, K. J., and S. B. Reed. 1999. "CPI Research Series Using Current Methods, 1978-98." Monthly Labor Review 122, no. 6 (June): 29-38.

Velde, F. 2006. "An Alternative Measure of Core Inflation." Federal Reserve Bank of Chicago Economic Perspectives 30, first quarter: 55-65.

Wynne, M. 1999. "Core Inflation: A Review of Some Conceptual Issues." Federal Reserve Bank of Dallas Working Paper no. 9903, June.

Timothy Clark, Astrid Dick, Beverly Hirtle, Kevin J. Stiroh, and Robard Williams

The Role of Retail Banking in the U.S. Banking Industry: Risk, Return, and Industry Structure

- In recent years, retail banking has become a key area of strategic emphasis in the U.S. banking industry, as evidenced by rising trends in retail loan and deposit shares on commercial bank balance sheets and a continuing increase in the number of bank branches.
- This "return to retail" contrasts with the 1990s, when banks sought to diversify revenues, deemphasize branch networks, and target financial services to a broader range of clients.
- An analysis of this strategic shift suggests that interest in retail banking fluctuates in predictable ways with the performance of nonretail banking and financial market activities.
- The recent "return to retail" episode may be more persistent than past cycles because it is being driven almost entirely by the very largest U.S. banks, which have been building large branch networks and investing in other retail banking infrastructure.

1. Introduction

The U.S. banking industry is experiencing renewed interest in retail banking. These activities—broadly defined as the range of products and services provided to consumers and small businesses—have grown in importance over the past several years. Retail-related positions now account for larger shares of commercial bank balance sheets, and the number of bank branches continues to grow. The recent focus on retail contrasts sharply with industry views held during the 1990s, when banks' attention turned to broadening products, diversifying revenues, substituting alternative delivery channels for branches, and offering a multitude of financial services to all types of retail, corporate, and wholesale customers.

This "return to retail" is reflected in a greater number of media reports on retail banking activities, in the frequency with which retail banking activities have been mentioned in banks' public statements, and in the attention given to these activities by industry analysts.[1] A 2004 report by Standard and Poor's—"Retail Sector Anchors Large Complex Banks in U.S."—and a

[1]For instance, a search of *American Banker* online indicates that 501 articles published between January 1, 2003, and December 31, 2004 (or 3.5 percent of all articles published during that period), included the phrase "retail banking," compared with 401 articles published between January 1, 1999, and December 31, 2000 (2.2 percent of all articles published).

Timothy Clark is an assistant vice president, Beverly Hirtle a senior vice president, and Robard Williams a vice president at the Federal Reserve Bank of New York; Astrid Dick was an economist and Kevin J. Stiroh a vice president at the Bank at the time this article was written.
Correspondence: <beverly.hirtle@ny.frb.org>

The authors thank two anonymous referees and Argia Sbordone for helpful comments. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

2003 Salomon Smith Barney discussion of U.S. banking becoming "refocused on retail" typify the view that retail has become a key area of strategic emphasis in the U.S. banking industry. Indeed, the renewed focus on retail activities seems to have been a key motivation behind a number of recent large-bank mergers, such as Bank of America's acquisition of FleetBoston Financial and JPMorgan Chase's acquisition of Bank One.[2]

This article documents the "return in retail" in the U.S. banking industry and offers some insight into why this shift has occurred. Trends in retail loan shares, retail deposit shares, the balance sheets of U.S. consumers, and the number of bank branches all indicate an increased focus on retail activities. We discuss the effect of this focus on individual banks and ask whether the related investment in infrastructure—principally, branch networks—is justified and sustainable for the industry as a whole. We examine a range of external sources: reports by equity analysts, rating agencies, and consulting firms; discussions and data provided by banking companies in annual reports, investor presentations, and other public outlets; and academic research examining various aspects of retail banking.

At the bank level, the principal attraction of retail banking seems to be the belief that its revenues are stable and thus can offset volatility in the nonretail businesses.

At the bank level, the principal attraction of retail banking seems to be the belief that its revenues are stable and thus can offset volatility in the nonretail businesses, such as corporate and commercial real estate lending, trading, and capital market activities. Some banking industry analysts go even further, claiming that retail banking offers high returns along with low risk. We present some evidence that retail banking activities offered high risk-adjusted returns relative to nonretail activities in the early 2000s, but that more recently the returns from retail and nonretail banking have converged. More formal analysis of large, publicly traded bank holding companies from 1997 to 2004 by Hirtle and Stiroh (forthcoming) suggests that both risk and return decline as these firms become more focused on retail banking activities. This finding, which is consistent with traditional finance theory, highlights the importance of taking a longer run perspective when considering how risk and return are affected by broad shifts in business strategy.

At the aggregate level, our review shows that interest in retail banking fluctuates in rather predictable ways with the performance of nonretail banking and financial market activities. We document the features that the recent "return to retail" has in common with past cycles, but also recognize important factors suggesting that this episode may be more persistent. In particular, this retail banking cycle is being driven almost entirely by the very largest U.S. banking firms. Branching deregulation in the 1990s enabled large banks to compete more effectively with smaller local institutions by establishing branch networks spanning large geographic areas. These banks have made substantial investments in large branch networks and other retail banking infrastructure, a development that seems unlikely to unwind quickly. Retail banking, for example, accounts for 50 to 75 percent of revenues at many large bank holding companies, so the key role of the very largest banks in the "return to retail" gives extra weight to these developments.

Our study proceeds as follows. Section 2 begins with an overview of retail banking and describes how its activities are managed at many large bank holding companies. We then examine, in Section 3, historical trends in retail banking and document the renewed interest in retail. In Section 4, we consider some of the factors that contributed to the most recent surge in retail activities. From a microeconomic perspective, we review claims by banks and industry analysts about risk and return, and ask whether the claims stand up to the available evidence. From a macroeconomic view, we investigate how the interest rate environment may have affected these observed trends. We also address the question of whether the recent emphasis on retail is likely to be permanent or transitory. Section 5 summarizes our findings and discusses areas of future research.

At the aggregate level, our review shows that interest in retail banking fluctuates in rather predictable ways with the performance of nonretail banking and financial market activities.

2. What Is Retail Banking?

Retail banking is the cluster of products and services that banks provide to consumers and small businesses through branches, the Internet, and other channels. As this definition implies, banks organize their retail activities along three complementary dimensions: customers served, products and services

[2]See, for example, *Wall Street Journal* (2003) and Deutsche Bank Securities (2004).

offered, and the delivery channels linking customers to products and services. (The box illustrates how several large banks describe their own retail banking activities.)

Organizationally, many large banking companies have a distinct "retail banking" business unit with its own management and financial reporting structure. Our description focuses on the common elements across these retail banking business segments. There are, however, differences in the way institutions organize and manage retail activities, so we also discuss the most significant variations.

Consumers and small businesses are typically the core retail banking customers. Consumers are served almost entirely by the retail banking business unit, although some large organizations have a separate subprime consumer finance

In Their Own Words: How Banks Describe Their Retail Banking Activities

These descriptions are from the 2005 annual reports of four large banks. This group certainly does not constitute an exhaustive list of institutions that provide detailed information on their retail banking activities. However, the passages cited here are representative of the information provided by large banking organizations that identify distinct retail business segments in their annual reports.

Bank of America

"Bank of America serves more than 38 million consumer and small business relationships in the nation's fastest-growing and most diverse communities. Sales, service, and fulfillment are provided through more than 5,800 banking centers and nearly 17,000 ATMs in 29 states and the District of Columbia. We also offer our customers the leading online banking service in the United States, with more active online bill payers than all competing banks combined, as well as a 24-hour telephone banking service that earns high ratings for speedy and easy self-service. With product and sales teams coordinating closely within these various distribution channels, Bank of America has grown to become the nation's largest provider of checking and savings services, the No. 1 credit and debit card provider (effective with completion of the MBNA merger on Jan. 1, 2006), the No. 1 small business lender, the leading home equity lender, and the fifth-largest originator of consumer mortgages."

Citigroup

"Citigroup's Global Consumer Group provides a wide array of banking, lending, insurance, and investment services through a network of 7,237 branches, 6,920 ATMs, 682 Automated Lending Machines (ALMs), the Internet, telephone and mail, and the Primerica Financial Services sales force. Global Consumer serves more than 200 million customer accounts, providing products and services to meet the financial needs of both individuals and small businesses."

JPMorgan Chase

"Retail Financial Services helps meet the financial needs of consumers and small businesses. We provide convenient consumer banking through the nation's second-largest ATM network and fourth-largest branch network. We are the second-largest home equity originator, the fourth-largest mortgage originator and servicer, the largest non-captive originator of automobile loans, and a top provider of loans for college students. We serve customers through more than 2,600 bank branches and 280 mortgage offices, and through relationships with 15,600 auto dealerships and 2,500 schools and universities. More than 11,000 branch salespeople assist customers with checking and savings accounts, mortgage and home equity loans, small business loans, investments, and insurance across our 17-state footprint from New York to Arizona. An additional 1,500 mortgage officers provide home loans throughout the country."

Wells Fargo and Co.

"The Community Banking Group offers a complete line of banking and diversified financial products and services to consumers and small businesses with annual sales generally up to $20 million in which the owner generally is the financial decision maker. Community Banking also offers investment management and other services to retail customers and high-net-worth individuals, insurance, securities brokerage through affiliates, and venture capital financing. These products and services include the *Wells Fargo Advantage Funds*[SM], a family of mutual funds, as well as personal trust and agency assets. Loan products include lines of credit, equity lines and loans, equipment and transportation (recreational vehicle and marine) loans, education loans, origination and purchase of residential mortgage loans, and servicing of mortgage loans and credit cards. Other credit products and financial services available to small businesses and their owners include receivables and inventory financing, equipment leases, real estate financing, Small Business Administration financing, venture capital financing, cash management, payroll services, retirement plans, Health Savings Accounts, and credit and debit card processing. Consumer and business deposit products include checking accounts, savings deposits, market rate accounts, Individual Retirement Accounts (IRAs), time deposits, and debit cards. Community Banking serves customers through a wide range of channels, which include traditional banking stores, in-store banking centers, business centers, and ATMs. Also, *Phone Bank*[SM] centers and the National Business Banking Center provide 24-hour telephone service. Online banking services include single sign-on to online banking, bill pay, and brokerage, as well as online banking for small business."

unit with its own brand identity. At the other end of the spectrum, services used primarily by high-net-worth individuals and households, such as trust and brokerage services, are nearly always provided by business units that specialize in these activities and offer them to all bank customers (for example, retail brokerage services are provided by a larger brokerage or asset management business segment).

The small businesses served by retail banking business units range from small start-ups and sole proprietorships to more established firms with annual revenues of $1 million or more. Most banks define "small business" by annual sales or revenue volume, generally with a cutoff separating small business customers and middle-market corporate customers. This cutoff can be anywhere between $1 million and $20 million in annual sales (larger banks tend to have larger cutoffs). At some banks, middle-market corporate customers—those with sales volumes up to $100 million to $250 million—are also served by the retail banking business unit, although it is increasingly common to serve these midsize businesses along with large corporate customers in a single corporate banking business line.

In terms of products and services, deposit taking is the core retail banking activity on the liability side . . . . On the asset side of the balance sheet, the key retail banking products are consumer credit and small business loans.

In terms of products and services, deposit taking is the core retail banking activity on the liability side. Deposit taking includes transaction deposits, such as checking and NOW accounts, and nontransaction deposits, such as savings accounts and time deposits (CDs). Many institutions cite the critical importance of deposits, especially consumer checking account deposits, in generating and maintaining a strong retail franchise. Retail deposits provide a low-cost, stable source of funds and are an important generator of fee income. Checking accounts are also viewed as pivotal because they serve as the anchor tying customers to the bank and allow cross-selling opportunities (Dick et al. 2006).

On the asset side of the balance sheet, the key retail banking products are consumer credit and small business loans. Consumer credit includes credit cards, mortgages, home equity lending, auto loans, education loans, and other personal loans. Some very large banking organizations have national consumer credit operations—principally for credit cards and mortgages, though also sometimes for auto loans—that are managed separately from the main retail banking business line. The separate management of these national businesses most likely reflects past regulatory restrictions against interstate banking and branching that, until the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994, precluded banks from operating branches on a national scale.

Although loans and deposits are the primary products, retail banking units provide a range of other financial services to consumers and small businesses. For individual consumers, these services include sales of investment products (such as mutual funds and annuities), insurance brokerage, and financial and retirement planning. For small businesses, they include merchant and payments services, cash handling, insurance brokerage, and payroll and employee benefits services.

Banks generally see the branch network as the central delivery channel in retail banking and perhaps the single most important component of the retail franchise. This view represents a significant turnaround from a decade ago, when branches were seen as an expensive and outmoded way to deliver retail banking services—one almost certain to be supplanted by remote delivery channels such as ATMs, telephone call centers, and the Internet.[3] These remote channels are now viewed as complements to the branch network. Call centers are used primarily for customer service and problem resolution, while online/electronic banking is used for information dissemination, transactions, and, increasingly, new-account origination. Finally, branches are pivotal for attracting new customers and generating cross-selling opportunities. Branches are now often staffed by licensed personnel who can sell investment products and insurance and who may also be linked to formerly stand-alone business lines, such as a mortgage or finance company (Dick et al. 2006).

The three dimensions of the retail banking business—customers, products and services, and the delivery channels linking customers with products—are interrelated.

[3]Orlow, Radecki, and Wenninger (1996) summarize the views on branching that prevailed in the mid-1990s.

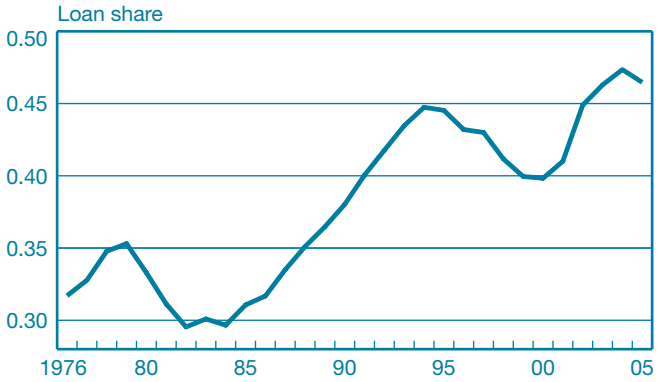
Clearly, the three dimensions of the retail banking business—customers, products and services, and the delivery channels linking customers with products—are interrelated. Consumers and small businesses constitute a coherent customer group largely because of commonalities in the financial products and services they use. These products and services have similar risk characteristics (both generate large pools of small, diversifiable loans where the primary risk is exposure to the business cycle), generating economies of scope in risk management. In some cases, consumers and small businesses use precisely the same products (credit cards are an important source of credit to both consumers and to the very smallest businesses). Furthermore, consumers and small businesses are both well served through the branch network. Finally, branches are the key retail banking delivery channel, largely because of the pivotal role they play in attracting and retaining consumer deposits, the core retail banking product. Thus, the three dimensions must be viewed together in order to understand retail banking completely.

3. The Evolution of Retail Banking

To gauge the evolving importance of retail banking, one would ideally examine a single, comprehensive measure of retail banking activity that could be calculated for individual banks and for the industry as a whole. Potential candidates might be the share of revenue or profit derived from retail activities or the share of risk capital allocated to these business units. Both measures are holistic in that they condense the full range of retail activities—both those that generate balance-sheet positions and those that do not—into a single measure that is comparable across business lines in the firm. Unfortunately, only a small number of large banks include in their annual reports and other public financial statements the figures on revenue, profits, and risk capital for identifiable retail business lines. Such information is not readily available for most banks.

To generate consistent measures of retail banking activity, we turn to an alternative source: data from regulatory reports. The advantage of using such data is that they are available on a consistent basis for all banks over a relatively long period.[4] We focus on three primary indicators of retail activity: retail lending (one-to-four-family mortgages, home equity lending, credit card loans, and other consumer loans), retail deposits (NOW accounts, savings accounts, and small time deposits), and the number of bank branches.[5] We also examine the share of household assets held as deposits.

Chart 1
Retail Loan Share
Credit Card, Other Consumer, and One-to-Four-Family Mortgage Loans as a Share of Total Loans



Source: Federal Financial Institutions Examination Council, Reports on Condition and Income.

Observed trends in retail loan shares, retail deposit shares, the balance sheets of U.S. consumers, and the number of bank branches all indicate an increased focus on retail activities. Chart 1 shows that for the U.S. banking system as a whole, the share of loans made to retail customers has increased significantly since the early 1980s, though with noticeable waves during this period. Much of the long-run increase is due to the growth of mortgage-related lending and, to a lesser extent, credit cards, particularly at larger institutions. This result reflects two developments: the decline, beginning in the mid-1980s, of the thrift industry, a traditional sector for mortgage lending, and technological changes that enabled large banks to realize scale economies in credit card and mortgage activities.[6]

The recent surge in retail banking is evident in the retail loan share, which has increased sharply since 2000. This increase has been led by growth in home equity lending and, to a somewhat

[4]We use balance-sheet data on loans and deposits from the Federal Financial Institutions Examination Council Reports on Condition and Income (the Call Reports) filed quarterly by all commercial banks (available at <http://www.chicagofed.org/economic_research_and_data/commercial_bank_data.cfm>), as well as data on branch ownership from the Summary of Deposits Reports commercial banks and thrifts file annually with the Federal Deposit Insurance Corporation (available at <http://www2.fdic.gov/sod>).

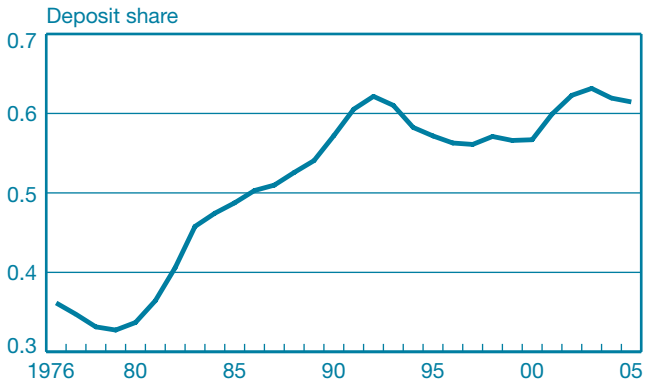
[5]Given the typical range of retail banking activities, small business loans should also be included in the retail loan share variable. Unfortunately, data on small business lending are available only starting in 1993, so we cannot construct a consistent historical sample. However, small business loans are a small share of overall loans held by U.S. banks (averaging 5 percent from 1993 to 2005), so the series omitting small business loans seems like a reasonable approximation. The correlation between the retail banking loan share, including and excluding small business loans, is 0.98 over 1993 to 2005, suggesting that this approximation is unlikely to have distorted the pattern depicted in Chart 1.

[6]For instance, Carter and McNulty (2005) find that large banks have an advantage in credit card lending, which the researchers attribute to technological innovation and reliance on "hard information" in lending.

Chart 2

Retail Deposit Share

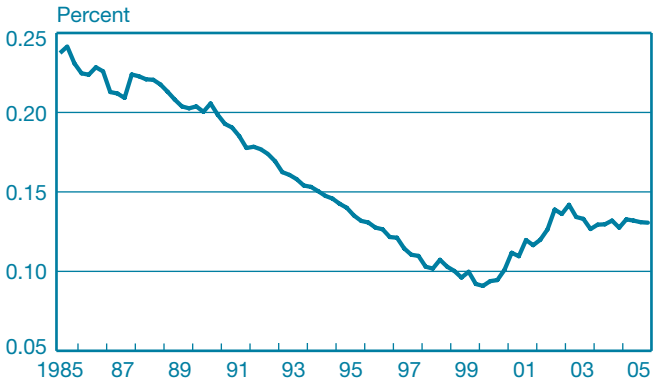
NOW, Savings, and Small Time Deposits as a Share of Total Deposits



Source: Federal Financial Institutions Examination Council, Reports on Condition and Income.

Chart 3

Share of Household Assets Held as Deposits



Source: Federal Reserve Statistical Release Z.1.

lesser extent, in credit card loans and one-to-four-family mortgages.[7] Chart 2 illustrates similar growth in retail deposits over this recent period, primarily reflecting a surge in savings account deposits.[8] The long-run growth of retail-related positions is also evident in the deposit data, which show retail-related deposit balances increasing during the 1980s with the removal of Regulation Q's ceilings on deposit interest rates. Both retail shares have cycled over time, however, showing similar peaks in the early to mid-1990s.

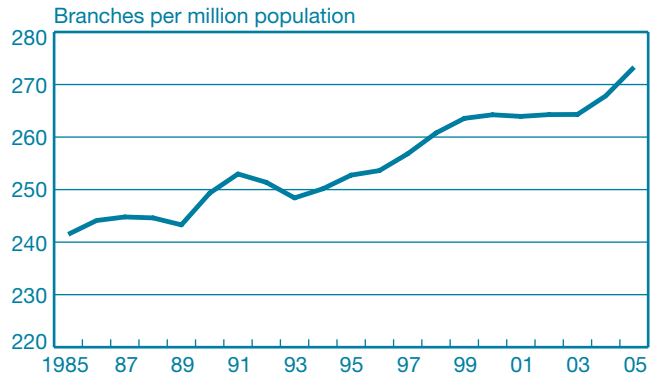
The growth of retail-related positions on banks' balance sheets is mirrored by corresponding growth in bank-related positions on the household balance sheet. Chart 3 illustrates the share of household assets held in the form of deposits.[9] Following years of steady decline, this ratio began to rebound after 2000, reaching levels comparable to those in the mid-1990s. Some part of this increase reflects a fall in the value of household assets attributable to the stock market's sharp decline in the early 2000s. Even so, household deposit growth accelerated over this period, and deposits as a share of household assets increased even after controlling for declining equity holdings. While households continue to hold a considerably smaller share of their assets in the form of bank deposits than was true in the 1980s, the recent upswing in this share is a marked departure from more than fifteen years of steady decline.

Along with growth in balance-sheet positions, the number of bank branches has been going up (Chart 4). Bank branches per capita have been increasing since the mid-1990s, and this growth has accelerated since 2003. Furthermore, an increasing portion of branches are held by a relatively small number of large banks. As of mid-2003, nearly 25 percent of U.S. branches were held by bank and thrift holding companies with 1,000 or more branches, up from 11 percent in 1994 (Hirtle and Metli 2004).

These four metrics of retail intensity show similar, but not identical, trends. For instance, during the early 1990s and in the

Chart 4

Bank Branches per Capita



Sources: Federal Deposit Insurance Corporation; U.S. Census Bureau.

[7]These figures reflect loans held on the balance sheet. Because significant portions of some types of retail lending—most notably, credit card loans and one-to-four-family mortgages—are securitized, the figures most likely understate the portion of loans originated to retail customers.

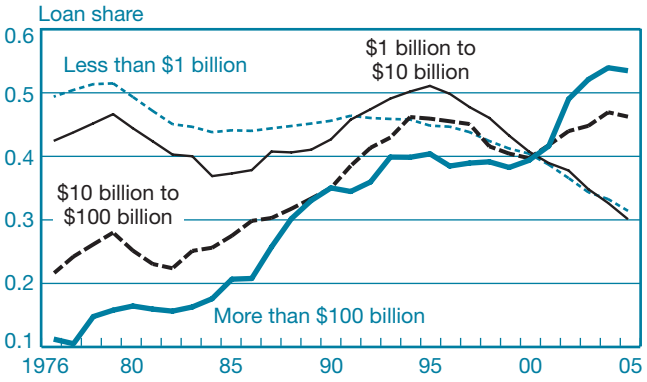
[8]We should note, however, that for the U.S. banking industry, deposits as a share of assets have been declining for several decades. This result reflects rising capital ratios, growing use of other borrowed funds, and increased issuance of subordinated debt.

[9]The ratio reported in Chart 3 is the share of assets of households and nonprofit organizations held in the form of currency, checkable deposits, and time and savings deposits as reported in the Federal Reserve's Flow of Funds Accounts. The deposit figures include deposits held at savings institutions and credit unions as well as at commercial banks.

Chart 5

Retail Loan Share by Bank Holding Company Asset Size

Credit Card, Other Consumer, and One-to-Four-Family Mortgage Loans as a Share of Total Loans

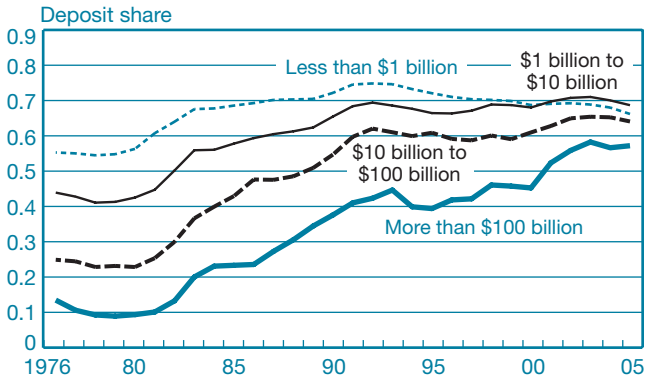


Source: Federal Financial Institutions Examination Council, Reports on Condition and Income.

Chart 6

Retail Deposit Share by Bank Holding Company Asset Size

NOW, Savings, and Small Time Deposits as a Share of Total Deposits



Source: Federal Financial Institutions Examination Council, Reports on Condition and Income.

current retail banking episode, the upswing in bank branches per capita begins well after the surge in retail loans and deposits. While this lag may simply reflect a longer reaction time for physical investments to come on line as compared with financial ones, it also points to the varied nature of retail banking and highlights the difficulty in creating a single measure that captures retail banking for all firms. Hirtle and Stiroh (forthcoming) address this issue by extracting the principal component of various measures of retail activity at the bank level; they show that the common factor declined in the late 1990s and then rose substantially after 1999. This finding supports our claim of an important shift toward retail activities in recent years.

Large banks have played an especially important role in the industry's renewed interest in retail banking. Charts 5 and 6 present the retail-related shares of loans and deposits for banks in different size cohorts based on total assets (deflated using the CPI and measured in 2004 dollars) between 1976 and 2005.[10] Over this long period, growth in retail-related loans was driven primarily by the larger banks, those most in a position to realize the economies of scale inherent in the mortgage and credit card business lines. In contrast, the retail deposit share increased for banks of all sizes, most likely reflecting the industrywide impact from the removal of ceilings on deposit interest rates in the early and mid-1980s.

The more recent growth in retail-related loan positions has been driven entirely by banks with assets exceeding $10 billion, especially the very largest in this group. The retail loan share at banks with assets exceeding $100 billion, for example, increased from 38 percent in 1999 to nearly 55 percent at the end of 2005 (Chart 5). In contrast, the retail loan share at smaller holding companies actually declined over the same period.[11] As a result, large banks now have a higher share of retail loans than do smaller banking firms.

Chart 6 shows a similar pattern for the retail deposit share in recent years. The overall increase has clearly been driven by the very largest banks, whose retail deposit share has grown steadily since the mid-1990s. In contrast, for smaller institutions over this period, the retail deposit share has trended slightly downward. Although these smaller institutions continue to have greater retail "intensity" by this measure, there has been a notable convergence across institutions of different asset sizes.

Consistent with these developments, retail banking is a significant source of revenue and profit for many large banking organizations. Data from a sample of large banks' annual reports and public financial statements suggest that between 50 and 75 percent of net operating revenue (net interest income plus noninterest income) is derived from retail banking activities at most of these institutions (Chart 7).[12] Table 1 presents similar information, obtained from a study by Citigroup Smith Barney, for a larger set of institutions in 2002.

[10] Assets for individual banks are aggregated so that the size cohorts are based on the assets of all banks within a holding company.

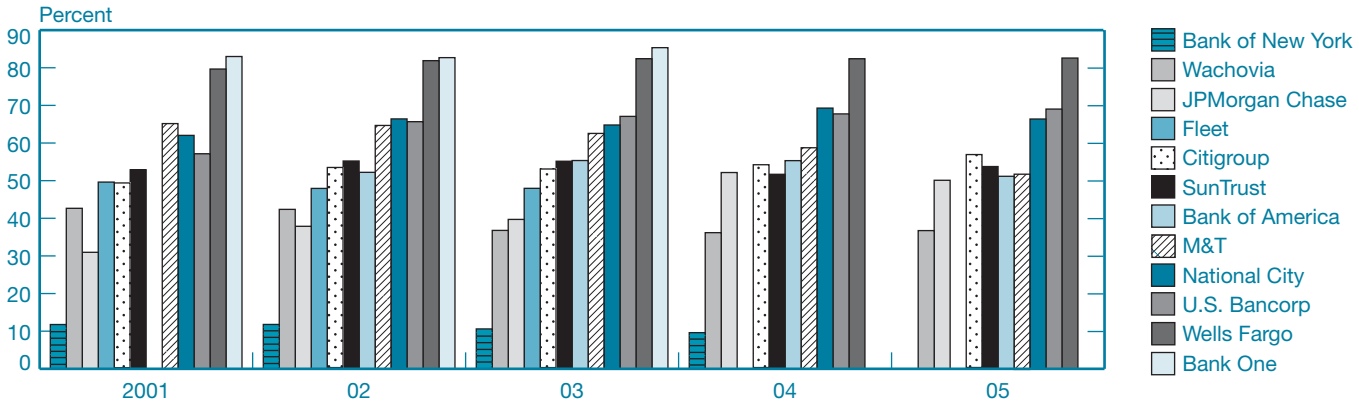
[11] These institutions are now holding higher shares of commercial real estate loans and construction and land development loans.

[12] The sample firms were selected based on asset size, branch network size, and whether they reported business segment financial information that allowed us to identify a retail banking business line consistent with our definition. This group does not necessarily represent an exhaustive list of U.S. banks for which such information may be available, but it is representative of a range of asset sizes and extent of retail focus among large banks.

Chart 7

Retail Revenue as a Share of Overall Revenue

Net Interest Income Plus Noninterest Income



Source: Bank holding company annual reports and quarterly earnings statements.

These data also suggest that retail activities account for 50 to 60 percent of revenue at a typical large bank.

The leading role of large banking organizations in the resurgence of retail banking is also reflected in the growth and redistribution of bank branches. As we observed, the number of bank branches has grown steadily since the early 1990s, and an increasing share is held by banking organizations with large branch networks. The historical consolidation of branches into these large networks has occurred primarily through mergers and branch purchases more than through de novo growth (Hirtle and Metli 2004).[13] This finding is consistent with the pattern in recent merger activity, much of which has focused on the expansion of banks' geographic footprints and reflects the new operating environment in a more deregulated era.

The consolidation of bank branches into very large branch networks can be linked to a combination of deregulation and technological change. A critical structural change in the U.S. banking industry over this period was the Riegle-Neal Act of 1994, which allowed nationwide branching and by 1997 had been adopted by virtually all states. This deregulation spurred a wave of industry consolidation that allowed banks to create the broader branch networks and increased branch penetration rates that are key to attracting new retail customers. In addition, it allowed banks to reap the benefits of technology-driven economies of scale. Dick (2006), for example, finds that deregulation in the 1990s significantly increased the size of branch networks, both in terms of the network's density in a given local market and in terms of the coverage over larger geographic areas.

These large networks, combined with innovations such as new credit-scoring technologies, have allowed large banks to compete more effectively with small community banks in the retail sector. Berger et al. (forthcoming), for example, find that large, multimarket banks were better able to compete against small banks in the 1990s, relative to the 1980s, presumably as a result of technological progress leading to an increase in scale economies in the management of larger organizations relative to smaller ones, and new lending technologies for small businesses that diminished the comparative advantage of small banks in servicing this segment.[14] Akhavein, Frame, and White (2005) find that large banks adopted new technologies for small business lending earlier than smaller banks did, using

The consolidation of bank branches into very large branch networks can be linked to a combination of deregulation and technological change.

[13]This is not to say that large banks did not create de novo branches. For instance, the 80 banking organizations with 100 or more branches in 2001 opened more than 2,100 de novo branches between June 2001 and June 2003. These same institutions, however, acquired 3,700 branches through mergers and purchases over this period, about two-thirds of gross branch expansion for these firms. These institutions closed or sold 3,700 branches, for net branch growth of approximately 2,100 over this period (Hirtle and Metli 2004).

[14]It is the large banks that expand geographically to multiple markets that enjoy these increases in efficiency, as opposed to large banks that only increase the scale of operations but remain in a single local banking market.

Table 1

Sources of Bank Revenue in 2002

Percent

	Retail Activities								
Bank	Consumer Banking	Credit Card	Total	Commercial Banking	Trust/Asset Management	Processing	Trading	Private Equity	Investment Banking
Charter One	87	1	88	12	1	0	0	0	0
TCF Financial	65	13	78	21	2	0	0	0	0
BB&T	76	0	76	19	4	1	0	0	0
National City	72	4	76	15	6	3	0	0	0
Bank One	39	33	72	14	11	0	3	0	0
Wells Fargo	67	3	70	21	10	0	0	0	0
Sovereign	66	1	67	31	2	0	0	0	0
National Commerce	65	2	67	26	1	5	0	0	0
Huntington	56	9	65	23	13	0	0	0	0
Amsouth	60	3	63	27	10	0	0	0	0
SunTrust	53	4	57	30	12	0	0	0	0
PNC	57	0	57	16	21	6	0	0	0
Regions	54	2	56	33	6	0	3	0	2
Union Planters	51	5	56	40	4	0	0	0	0
Bank of America	47	8	55	19	5	4	11	0	5
First Tennessee	50	4	54	22	7	2	16	0	0
SouthTrust	50	4	54	41	4	0	0	0	0
U.S. Bancorp	37	12	49	34	11	6	0	0	0
Fifth Third	46	2	48	37	8	8	0	0	0
FleetBoston	38	10	48	38	9	0	5	0	0
North Fork	47	0	47	52	1	0	0	0	0
Wachovia	47	0	47	29	14	0	1	0	10
Key	43	0	43	40	10	0	3	0	4
M&T Bank	43	0	43	50	7	0	0	0	0
Synovus	38	1	39	31	3	27	0	0	0
JPMorgan Chase	26	12	38	15	7	12	20	0	9
Comerica	34	1	35	65	0	0	0	0	0
Bank of New York	11	0	11	21	7	54	7	0	0
Mellon	0	0	0	30	46	16	8	0	0
State Street	0	0	0	6	9	64	21	0	0
Average	47	4	52	29	8	7	3	0	1
Median	49	2	55	28	7	0	0	0	0

Source: Citigroup Smith Barney (2003).

them especially to expand their lending to relatively opaque small businesses, a segment that had been traditionally dominated by small banks.

Research also suggests that branches held within very large networks—those with 1,000 or more branches—outperform those held in midsize networks (100 to 500 branches) in terms of generating higher deposits per branch (Hirtle forthcoming). These results are certainly consistent with the observed consolidation of branches into the large branch networks of multistate banking organizations. From the bank's perspective, the development of a branch network can be particularly valuable as a barrier to entry for potential competitors, as a form of advertising to attract consumers, and as a funding source to generate stable deposits. For instance, Dick (2007) finds that the leading banks in a market make larger investments in branch networks and that these investments grow with market population. An implication of this finding is that when profit opportunities arise in a market, such as those created by an inflow of new customers, large banks are likely to open new branches as a way to take up the additional demand and prevent further entry.

4. Understanding the "Return to Retail"

Having documented the increased focus on retail activities measured by bank assets, liabilities, and branch infrastructure, we now discuss some of the underlying forces that contributed to this strategic shift. We provide a brief overview of the perspective of industry analysts and summarize the issues most frequently raised. Our study then considers the microeconomic forces reflecting changes in risk and return opportunities as well as several macroeconomic factors associated with deregulation and aggregate conditions. Finally, we offer some discussion and speculation on whether the most recent shift toward retail is likely to be temporary or permanent.

4.1 Perspective from the Banking Industry

A long, consistent industry perspective is provided by *BusinessWeek*, which has produced an annual analysis of banking industry trends since the mid-1980s. We complement this perspective with commentary by industry analysts at investment banks, consulting firms, and rating agencies (Moody's and Standard and Poor's). This information from industry participants provides a useful perspective on the evolving perception of retail banking. It is somewhat "soft" information, however, so we also examine data on acquisitions and mergers by banking organizations, which support the trends identified in the industry reviews.

The mid-1990s . . . marked the emergence of the "new retail" model that emphasized alternative retail delivery channels such as telephone call centers, ATMs, and electronic delivery through the emerging Internet.

In the mid-1980s, the *BusinessWeek* reviews focused on the need for banks to generate new revenue streams and enter new markets (for example, securities, investment banking, and insurance) as a way to counter the negative effect of increased competition and disintermediation within traditional retail banking. This was the period in which households began to turn away from the banking sector and toward mutual funds, which were increasingly competing with banks for household assets (Chart 3). However, in the late 1980s, as the banking industry recovered from significant problems, *BusinessWeek* noted a renewed interest in retail activities and quoted a senior Citibank executive as saying that the view that investment banking would rescue banks is "seriously in question" (*BusinessWeek* 1988).

A focus on alternative sources of revenue returned in the early 1990s, with an emphasis on the need for regulatory changes that would allow banks to diversify further into securities and underwriting. This diversification can be seen in the growth in Section 20 subsidiaries, which allowed banks to underwrite corporate debt and equity issues. By 1994, *BusinessWeek* was reporting that noninterest income represented 40 percent of major banks' revenues, a sharp increase from previous levels, as those institutions attempted to diversify revenue streams. Stiroh (2004) documents a similar trend in noninterest income at large banking companies beginning in the early 1990s.

A shift in the strategic focus can be seen in the acquisition trends of large U.S. banks over the past decade, which echo the ebb and flow of interest in retail banking.

The mid-1990s, however, marked the emergence of the "new retail" model that emphasized alternative retail delivery channels such as telephone call centers, ATMs, and electronic delivery through the emerging Internet. At the time, a Chase Manhattan executive discussed the possibility of branches being supplanted by videoconferencing kiosks (*BusinessWeek* 1994), and the CEO of First Union predicted that customers would move away from branches as technology improved (*BusinessWeek* 1996). Similarly, Orlow, Radecki, and Wenninger (1996) quote executives of two major banks as saying they did not expect their institutions "to ever build another traditional branch." Electronic delivery, in particular, was seen as a low-cost alternative to high-cost branches. For instance, Moody's Investors Service (1996) lauded Bank of America's plan to "creatively destroy" its branch network and replace it with call centers, self-service ATMs, and supermarket locations. This focus on electronic banking continued through the late 1990s with the introduction of Internet-only bank operations, such as Bank One's Wingspanbank.com, a subsidiary that opened in June 1999.

Table 2

Bank Merger Announcements by Target Industry

Year	Depository Institutions	Nondepository Credit Institutions	Security and Commodity Brokers, Dealers, Exchanges, and Services	Insurance Carriers, Agents, Brokers, and Service	Holding and Other Investment Offices	Services and Other	Total
1994	188	14	3	4	5	1	215
1995	214	7	6	1	5	5	238
1996	158	5	3	3	7	3	179
1997	141	9	9	1	5	5	170
1998	151	14	11	3	12	1	192
1999	112	17	8	8	8	12	165
2000	95	12	8	13	17	19	164
2001	102	8	8	14	22	14	168
2002	62	6	25	21	14	11	139
2003	92	5	12	13	1	7	130
2004	92	13	13	9	7	11	145
Total	1,407	110	106	90	103	89	1,905

Source: Securities Data Corporation Mergers & Acquisitions Database.

Notes: The year is the date that the deal was announced. Depository Institutions are SIC 60; Nondepository Credit Institutions are SIC 61; Security and Commodity Brokers, Dealers, Exchanges, and Services are SIC 62; Insurance Carriers, Agents, Brokers, and Service are SIC 63 or 64; Holding and Other Investment Offices are SIC 67; Other includes SIC 20, 30, 40, 50, 65, 70, and 80.

By 2001, however, a pure-play Internet bank was viewed as a failed business model, and only a few experienced even modest success (Moody's Investors Service 2001). Interest in fee income and capital-market-related revenue sources surged in the late 1990s and early 2000s, while *BusinessWeek* (2001) quoted industry analysts who argued that traditional consumer deposit products were "dinosaurs."

The focus on capital market activities was short-lived, however. By 2002, the U.S. economy had experienced the bursting of the NASDAQ bubble, the events of September 11, and a massive decline in investment banking activities. Given a growing realization of the risks associated with capital market activities (volatility in trading revenue, the reputational effects of the corporate governance scandals, and the resultant compliance costs associated with regulatory reform such as the Sarbanes-Oxley Act of 2002) along with awareness of the operational difficulties associated with the diversified model (such as culture clashes between commercial and investment banking), interest in retail banking once again emerged, this time with a renewed emphasis on branches. Moody's Investors Service (2004), for instance, highlights the shift toward de novo branching and emphasizes the sharp reversal in strategy from the earlier period when banks were embracing alternative distribution channels.

The renewed interest in retail is also apparent in the mergers of First Union/Wachovia, Citigroup/Golden State, Bank of America/FleetBoston Financial, and JPMorgan Chase/Bank One, which were all motivated in large part by retail concerns. The Bank of America deal, for example, was driven by the potential growth and geographic expansion of the branch network (*Wall Street Journal* 2003), while the JPMorgan deal highlighted the stability of retail activities (Deutsche Bank Securities 2004). Similarly, Citigroup's sale of its Travelers Life and Annuity business to MetLife was viewed as part of a larger strategy to renew focus on consumer banking and abandon the financial supermarket model (*American Banker* 2005; *BusinessWeek* 2005). In general, the discussion of the motivations behind the recent mergers of large banks is very different from the discussion around the large deals, many involving nonbanks, in the 1990s.

Finally, a shift in the strategic focus can be seen in the acquisition trends of large U.S. banks over the past decade, which echo the ebb and flow of interest in retail banking. Table 2 summarizes acquisition trends by large U.S. banking organizations from 1994 to 2004 as reported by Securities Data Corporation.[15] The data show trends in bank merger and acquisition activity that correspond to the forces described

[15]Bank acquisitions are broken down into acquisitions of other depository institutions; nondepository credit institutions; security and commodity brokers, dealers, exchanges, and services; insurance carriers, agents, brokers, and service; holding and other investment offices; and all other. While one would also be interested in the size of deals over time, the values of all deals were not available, so Table 2 focuses on the number of announced deals.

earlier. For example, banks' acquisitions of other banks peaked in the mid-1990s after the Riegle-Neal Act of 1994 removed restrictions on interstate banking and branching. Bank acquisitions of nonbank subsidiaries surged after the Gramm-Leach-Bliley Act of 1999 liberalized activity restrictions in 1999 and interest in retail waned. Acquisitions of securities and brokerage firms and of insurance firms, for example, peaked in 2002, while acquisitions of holding and other investment offices peaked in 2001. The belief in the benefits of diversified financial firms and nontraditional activities during the late 1990s and early 2000s is manifest in the rise of nonbank acquisitions, while the return to a focus on core banking operations is seen in the mergers of large banks in recent years.

4.2 Microeconomic Factors

Banking industry analysts and the banks themselves consistently point to the stability of revenue and profit as the most important feature of retail banking and a key motivation for the recent interest. In particular, retail stability is seen as valuable for large banks seeking to offset the volatility of riskier business lines, such as trading and other capital-market-related activities. Recent discussions of retail activities in large banks' annual reports, analyst presentations, and press releases highlight retail as a core source of stable, predictable earnings

[The] stability of retail-related activities is typically attributed to several factors. The most important one is that retail banking is fundamentally a consumer-based business.

in times when other sources of revenue have been comparatively weak. For example, Standard and Poor's (2004) identifies retail banking as "an island of stability in the last cycle," while Moody's (2003) highlights the "low correlation to the lending business, creating earnings diversity" as a key benefit from retail activities. Standard and Poor's (2004) also points to the relative volatility associated with nonretail activities such as large corporate lending, investment banking, and emerging-market activities.

This stability of retail-related activities is typically attributed to several factors. The most important one is that retail banking is fundamentally a consumer-based business. The resilience of the consumer sector in recent years has almost certainly contributed to the stability of retail banking. An important corollary of this observation is that retail banking will likely be a stable and growing business only as long as the consumer sector remains strong and stable.

A second important factor in the stability of retail banking is that it serves a large number of small customers. The granular nature of the retail lending portfolio—which contains a large number of small, often collateralized loans—means that the lending income may be less volatile over time because of diversification across customers. In essence, the retail lending

A second important factor in the stability of retail banking is that it serves a large number of small customers.

portfolio is exposed primarily to cyclical or macroeconomic risk, rather than to borrower-specific exposures (concentration risk). This is one specific example of how retail banking stability relies on the continued strength of the consumer sector.

Finally, some part of the stability in retail banking revenues may reflect natural hedges within retail banking—in other words, products or services within the business that respond differently as market conditions change. One example cited by bankers is the low or negative correlation between mortgage originations and deposit margins. Deposit margins—the difference between rates paid on retail deposits and alternative market funding rates such as the federal funds rate—are an important source of income in retail banking. In periods of low interest rates, deposit margins tend to be low, reducing the implicit income earned on deposit balances. Low rates, however, spur mortgage refinancing, which boosts fee income.[16] Changes in income flows from the two activities thus tend to offset one another over the interest rate cycle, giving greater stability to overall retail banking revenues.

This view on the relatively stable nature of retail banking is consistent with the academic literature and analysts' reports. Stiroh (2004), for example, shows that in the period from 1984 to 2001, noninterest income, particularly trading, fees, and other noninterest income, was more volatile than deposit service charges or net interest income. DeYoung and Rice (2004) show that "traditional" and "community" banks (defined as those with relatively high core deposit ratios) have a relatively low volatility of revenue, as do "nontraditional" banks (which include a wide range of large banks).[17]

Despite considerable evidence supporting the stability of retail, the evidence on the returns from retail banking activities is more mixed. A recent study by Morgan Stanley and Mercer

[16] See Dick et al. (2006) for a discussion of the views of retail bankers.

Oliver Wyman (2004), for instance, describes retail banking as the "Cinderella" of U.S. financial services, offering "high margins, stable income, and modest capital consumption." An important conclusion of the study is that retail-focused banks have offered higher risk-adjusted returns, particularly in recent years when wholesale portfolios were negatively affected by the recession and other macroeconomic developments.

Recent data from a set of large bank holding companies suggest that retail activities have offered high risk-adjusted returns relative to other business lines. Chart 8 reports the ratio of return on risk-adjusted equity (ROE) in retail business lines to ROE on nonretail activities for a small set of banks that report business-line-level returns in public financial statements.[18] Returns on retail activities consistently exceed those on nonretail activities, often by a margin of two-to-one or three-to-one, a finding consistent with claims that retail banking offers high returns relative to risk. Significantly, however, there also appears to be a cyclical element at play. The retail-to-nonretail ratio has declined since 2002 as returns in nonretail business lines recovered from relatively low levels during and just following the 2001 recession and subsequent capital market slowdown. This result suggests that it is important to use a relatively long-run perspective when considering how risk and return are affected by broad shifts in business strategy.

Some part of the stability in retail banking revenues may reflect natural hedges within retail banking—in other words, products or services within the business that respond differently as market conditions change.

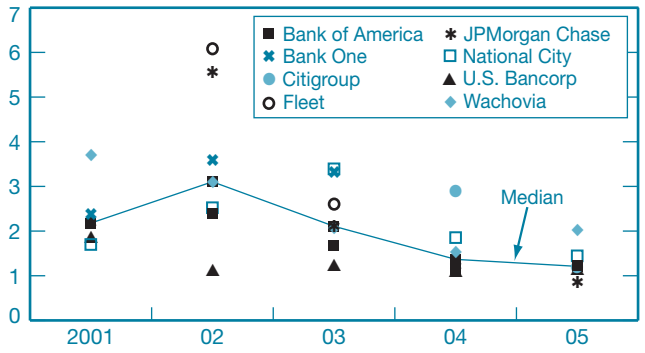
[17]Although the consumer loan portfolio as a whole may be granular and well collateralized, there are significant differences across different loan types in loss volatility over the business cycle (credit card loans, for instance, are uncollateralized and thus potential losses are higher). Thus, depending on a given bank's business mix, the stability of retail revenues will also vary.

[18]The ROE figures are not calculated consistently across bank holding companies. Most reflect returns on some form of risk-adjusted capital allocated to business lines, but both the risk capital calculations and the methods for allocating capital across business lines differ significantly across holding companies and sometimes over time for the same firm. That said, to the extent that ROE is calculated consistently across business lines within a holding company for a given year, the *ratios* of ROEs in different business lines should be reasonably comparable across institutions. Nonretail business lines exclude any returns or capital allocated to "corporate groups" or "parent" segments, since these tend to be cost centers rather than operational areas.

Chart 8

Ratio of ROE on Retail Business Lines to ROE on Nonretail Business Lines

For Large Bank Holding Companies Reporting ROE by Business Line



Source: Bank holding company annual reports and quarterly earnings statements.

Note: ROE is return on risk-adjusted equity.

A broader analysis by Hirtle and Stiroh (forthcoming) is consistent with the banks' perception that retail activities are generally stable, but it is less consistent with the notion that they are also high-return activities. That study compares both equity market returns and equity market volatility, a standard measure of risk, to various measures of retail intensity (the retail loan share, the deposit retail share, and branches per dollar of assets) for a sample of more than 700 bank holding companies from 1997 to 2004.

The results indicate that greater retail banking intensity is associated with lower equity market volatility for the very largest banks (those with assets greater than $10 billion). For small and midsize banks, the relationship between retail banking intensity and market volatility is weak. A key factor in this result is the role of branches: Greater branching intensity leads to lower volatility for large organizations, but to higher volatility for smaller ones. Regardless of organization size, however, higher retail banking intensity is associated with lower average returns based on both market and accounting data. Hirtle and Stiroh conclude that while retail banking may be a relatively stable activity, it is also a relatively low-return one.

Taken together, these findings offer mixed views on whether retail banking offers unusually attractive risk and return opportunities. Among industry observers, there seems to be a consensus, supported by the academic evidence, that retail activities tend to be more stable than other banking activities. In terms of returns, however, the evidence is much less

compelling for the notion that retail banking offers relatively high returns along with increased stability. To the extent that this belief motivated strategic shifts toward retail, one might conclude that these moves were less than fully justified.

4.3 Macroeconomic Factors

The earlier review suggests that interest in retail banking does cycle in relatively predictable ways with the performance of nonretail banking and financial market activities. Another perspective on this cyclicality comes from comparing the intensity of retail activities with cyclical movements in interest rates. Although not a rigorous analysis, it is useful to examine how broad measures of retail activity move with the interest rate environment.

Charts 9 and 10 illustrate the relationship between the retail shares of loans and deposits and changes in the slope of the yield curve. To highlight this relationship, we present the retail loan and deposit shares after removing trends.[19] The yield curve is measured as the difference between the annual average ten-year and one-year Treasury rates. The detrended retail loan share and retail deposit share both tend to increase as the yield curve steepens and to decline as the yield curve flattens. In the case of the retail deposit share, this movement is synchronous, with peaks in the yield curve corresponding to peaks in retail deposit share throughout the thirty-year sample period.[20] The retail loan share, in contrast, continues to rise for two to three years after the yield curve peaks, and then it trends down. We see the positive link between retail loans and the yield curve for three of the four yield curve peaks during our thirty-year sample period (mid-1970s, mid-1990s, and early 2000s). The link did not hold during the mid-1980s yield curve cycle, which may reflect the deep recession that preceded it.[21] Given this historical context, the run-up in the retail-related positions since 2000 does not seem out of proportion to earlier episodes, given comparable changes in the slope of the yield curve.

There are several possible explanations for the apparent positive link between retail loan and deposit shares and the yield curve. One explanation has to do with the attractiveness

[19]The trend is removed by regressing the retail ratio on a quadratic time trend (time and time-squared) and using the residuals as the detrended series. The results are similar if a simple linear trend is used.

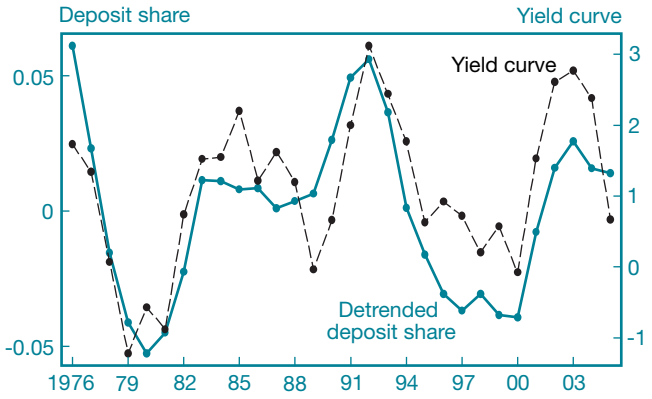
[20]The positive relationship between the retail deposit share and the slope of the yield curve holds when the sample is split by bank holding company asset size. The retail deposit share moves in sync with the yield curve for all four size cohorts examined (under $1 billion, $1 billion to $10 billion, $10 billion to $100 billion, and more than $100 billion in assets).

[21]In contrast to the results for the retail deposit share, the positive correlation between retail loans and the yield curve is significant only for the large banking companies.

Chart 9

Retail Deposit Share and the Yield Curve

NOW, Savings, and Small Time Deposits as a Share of Total Deposits



Sources: Federal Financial Institutions Examination Council, Reports on Condition and Income; Federal Reserve Statistical Release H.15; authors' calculations.

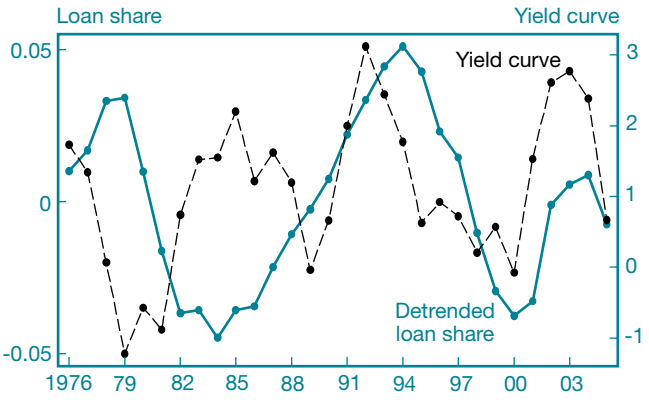
Note: Retail deposit share is detrended using a quadratic time trend.

of the "carry trade" when the yield curve is steep. When longer term rates are significantly higher than shorter term ones, banks may actively increase relatively low-cost retail deposits to fund longer term retail loans such as mortgages. Alternatively, to the extent that interest rate and yield curve movements affect profitability in nonretail banking and financial market activities, the relationship between retail banking and the yield curve may simply be another manifestation of the retail/

Chart 10

Retail Loan Share and the Yield Curve

Credit Card, Other Consumer, and One-to-Four-Family Mortgage Loans as a Share of Total Loans



Sources: Federal Financial Institutions Examination Council, Reports on Condition and Income; Federal Reserve Statistical Release H.15; authors' calculations.

Note: Retail loan share is detrended using a quadratic time trend.

nonretail cycle discussed above. Whatever the explanation, the key point is that the recent emphasis on retail banking seems consistent with historical experience, both in terms of the last significant steepening of the yield curve in the early 1990s and in terms of a broader pattern relating to profitability in other banking sectors.

4.4 Implications for the Future

The recent surge of interest in retail banking raises a number of questions for the future structure and performance of the U.S. banking industry. Perhaps the most significant is whether the widespread focus on retail banking and the related investment in infrastructure—principally, branches—is justified and sustainable. Is there anything to suggest that this upswing in retail banking focus will be more permanent

The recent surge of interest in retail banking raises a number of questions for the future structure and performance of the U.S. banking industry. Perhaps the most significant is whether the widespread focus on retail banking and the related investment in infrastructure—principally, branches—is justified and sustainable.

than those in the recent past? If so, what do these developments imply for the risks and future performance of the U.S. banking system?

Our analysis of industry trends, business commentaries, and studies of the effect of deregulation suggest that the rising focus on retail activities is not unprecedented and can be more accurately characterized as a "return to retail." As in past episodes, this "return to retail" undoubtedly reflects a number of transitory forces, including the evolving relative performance of different banking industry activities. Some part of the recent surge in interest has almost certainly been driven by volatility in capital markets and wholesale banking activities, raising the possibility that interest in retail banking will abate when these activities perform more strongly.

At the same time, more permanent factors such as deregulation and technology have clearly played an important role, suggesting that this latest episode may be more persistent than in the past. Deregulation has allowed large banks, for example, to lead the current focus on retail banking. Partly facilitated by branching deregulation in the 1990s, these banks have been accumulating retail-related assets and liabilities and have constructed large, geographically diverse branch networks—a development that seems unlikely to unwind

The key role of the very largest banks in the "return to retail" . . . suggests that the current episode may be more persistent, or have a longer lasting effect, than recent previous waves of interest in retail banking.

quickly. As the very largest banks continue to expand and consolidate in a deregulated banking environment, retail banking has become more concentrated among the largest banks, which, in turn, are relying to a greater extent on retail banking as a source of both revenue and stability. This represents an important shift in the strategic focus of many of the largest U.S. banking institutions.

The key role of the very largest banks in the "return to retail" gives extra weight to these developments and suggests that the current episode may be more persistent, or have a longer lasting effect, than recent previous waves of interest in retail banking. Furthermore, given the systemic importance of these institutions in the financial system, it is useful to think about the possible vulnerabilities that this shift may introduce.

The most obvious exposure is the implicit reliance on the growth and stability of the consumer sector. As noted above, retail banking will likely be a stable and growing business only as long as the consumer sector remains strong and stable. At a macroeconomic level, the consumer sector has indeed been quite resilient in recent years, but the focus on retail activities exposes banks to a slowdown in consumer spending brought on, for example, by a recession related to higher oil prices, falling real estate prices, or other adverse developments. Indeed, greater retail banking competition may itself have increased the likely severity of a consumer sector downturn if this competition has resulted in increased consumer debt levels.

A consumer sector shock could hurt the retail franchise on both the asset side (if defaults increase) and on the liability side (if retail deposits stop growing). To a large extent, consumer sector risk is inherent to any consumer-based business, and there are few steps banks might take to mitigate this exposure. The particular concern around retail banking, however, is that the very largest, systemically important banking institutions have been leading the surge in retail banking activity and thus may have taken on consumer sector "tail risk" to a much greater extent than in the past. Furthermore, because the emphasis on retail activities is fairly widespread among the largest firms, this "tail risk" exposure is a common risk to which they are simultaneously exposed. The banking system itself may thus be more exposed to a downturn in the consumer sector.

From a macroeconomic view, a deep and sustained downturn in the consumer sector would obviously have larger implications for the state of the U.S. economy, which would affect monetary and fiscal policy. To the extent that banks play a special role in the economy—given their ability to provide credit to informationally opaque small businesses—and that retail banking is particularly affected, a downturn in the consumer sector could impact the extent of a monetary policy response.[22]

5. Conclusion

Since around 2000, the U.S. banking industry has experienced a renewed interest in retail banking. Although there have been other periods in the past few decades when retail banking has been an important area of strategic focus, the recent cycle is particularly significant because of the role of the very largest banks. Many of these banks have been building large branch networks and increasing the share of retail-related positions on their balance sheets. As this article observes, retail banking is clearly an important source of revenue and profit for these firms and, given their systemic importance, it is important to understand the effect of this strategic focus not only for individual firms but for the banking system as a whole.

An interesting area for future research would be a deeper analysis of the macroeconomic factors driving the retail cycle. We present some suggestive evidence—for example, that U.S. banks' retail exposure varies with the interest rate cycle—and it would be useful to analyze this evidence more formally. A second area of potentially interesting work surrounds the question of cross-selling and relationship banking at the microeconomic level. Many industry observers have raised the possibility that retail banking offers important lock-in effects and motivates increased revenue from other business lines. This article does not address that issue empirically, and it remains an interesting item on the research agenda. A final theme for future analysis is to examine how the "return to retail" ultimately plays out. We have observed a range of expansion strategies: Some banks are growing their retail activities through mergers and acquisitions while others are focusing on de novo expansion. At this point, it is not clear which strategy, if any, dominates, and we will have to wait for new research over the next retail cycle.

[22]Bernanke (1983) highlights the role banks play in amplifying shocks, and Ashcraft (2005) documents banking's importance to local economic activity, particularly in terms of commercial and industrial lending and commitments.

References

Akhavein, J., W. S. Frame, and L. J. White. 2005. "The Diffusion of Financial Innovations: An Examination of the Adoption of Small Business Credit Scoring by Large Banking Organizations." Journal of Business 78, no. 2 (March): 577-96.

American Banker. 2005. "Citi Under Prince: Lots of Capital to 'Redeploy.'" February 2.

Ashcraft, A. B. 2005. "Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks." American Economic Review 95, no. 5 (December): 1712-30.

Berger, A. N., A. A. Dick, L. G. Goldberg, and L. J. White. Forthcoming. "Competition from Large, Multimarket Firms and the Performance of Small, Single-Market Firms: Evidence from the Banking Industry." Journal of Money, Credit, and Banking.

Bernanke, B. S. 1983. "Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression." American Economic Review 73, no. 3 (June): 257-76.

BusinessWeek. 1988. "Banks' Vaults Are Stuffed—With Bad Tidings." January 11.

———. 1994. "Another Year in 'Bank Heaven'?" January 10.

———. 1996. "Looking for Safety in Numbers." January 8.

———. 2001. "Banking and Securities." January 8.

———. 2005. "Citi: A Whole New Playbook." February 14.

Carter, D. A., and J. E. McNulty. 2005. "Deregulation, Technological Change, and the Business-Lending Performance of Large and Small Banks." Journal of Banking and Finance 29, no. 5 (May): 1113-30.

Citigroup Smith Barney. 2003. "Breaking Down Banks' Business Mix to Est. LTG Rates." April 25.

———. 2004. "Bank Weekly: Understanding the Impact of De Novo Expansion." March 12.

Deutsche Bank Securities. 2004. "Bank One Sells Out to JPMorgan in Strategic Merger: Initial Thoughts." January 15.

DeYoung, R., and T. Rice. 2004. "How Do Banks Make Money? A Variety of Business Strategies." Federal Reserve Bank of Chicago Economic Perspectives 28, fourth quarter: 52-68.

Dick, A. A. 2006. "Nationwide Branching and Its Impact on Market Structure, Quality, and Bank Performance." Journal of Business 79, no. 2 (March): 567-92.

———. 2007. "Market Size, Service Quality, and Competition in Banking." Journal of Money, Credit, and Banking 39, no. 1 (February): 49-81.

Dick, A. A., T. Clark, B. Hirtle, K. Stiroh, and R. Williams. 2006. "Profit Drivers in Retail Banking: The Bankers' Perspective." Unpublished paper, June.

Hirtle, B. J. Forthcoming. "The Impact of Network Size on Bank Branch Performance." Journal of Banking and Finance.

Hirtle, B. J., and C. Metli. 2004. "The Evolution of U.S. Bank Branch Networks: Growth, Consolidation, and Strategy." Federal Reserve Bank of New York Current Issues in Economics and Finance 10, no. 8 (July).

Hirtle, B. J., and K. J. Stiroh. Forthcoming. "The Return to Retail and the Performance of U.S. Banks." Journal of Banking and Finance.

Moody's Investors Service. 1996. "U.S. Money Center Banks: Strategic Choices Defining the Future." March.

———. 2001. "Banking and the Internet: An Update." July.

———. 2003. "Moody's Remains Upbeat on U.S. Banks' Ratings." June.

———. 2004. "Banks' De Novo Branch Expansion Strategies: Limited Ratings Impact." August.

Morgan Stanley and Mercer Oliver Wyman. 2004. "U.S. Retail Banking and Consumer Credit: An Agenda for Growth." June.

Orlow, D. K., L. J. Radecki, and J. Wenninger. 1996. "Ongoing Restructuring of Retail Banking." Federal Reserve Bank of New York Research Paper no. 9634, November.

References (Continued)

Salomon Smith Barney. 2003. "A Look at Banks' De Novo Expansion Plans." March 9.

Standard and Poor's. 2004. "Retail Sector Anchors Large Complex Banks in U.S." October 4.

Stiroh, K. J. 2004. "Diversification in Banking: Is Noninterest Income the Answer?" Journal of Money, Credit, and Banking 36, no. 5 (October): 853-82.

Wall Street Journal. 2003. "Branching Out: Bank of America Bets on Consumer." October 28.

UNDERSTANDING RISK MANAGEMENT IN EMERGING RETAIL PAYMENTS

Michele Braun, James McAndrews, William Roberds, and Richard Sullivan

New technologies used in payment methods can reduce risk, but they can also lead to new risks. Emerging retail payments are prone to operational and fraud risks, especially security breaches and potential use in illicit transactions. This article describes an economic framework for understanding risk control in retail payments. Risk control is a special type of *good* because it can protect one payment participant without diminishing the protection of other participants. As a result, the authors' economic framework emphasizes risk containment, primarily through the establishment and enforcement of risk management policies. Application of the framework to three types of emerging payments suggests that a payments system can successfully manage risk if it quickly recognizes problems, encourages commitment from all participants to control risk, and uses an appropriate mix of market and public policy mechanisms to align risk management incentives. The authors conclude that providers of emerging payment methods must mitigate risk effectively or face rejection in the payment market.

EPR Executive Summary available

EPR EXECUTIVE SUMMARIES

Visit our website for concise summaries of *Economic Policy Review* articles.

Our online publication *EPR Executive Summaries* condenses many of the articles published in the *Review*. Readers of the summaries will find timely, policy-oriented studies that are easy to absorb.

The summaries make the technical research of New York Fed economists more accessible to policymakers, educators, business and financial leaders, and others. The series is designed to foster a fuller understanding of our research among those who are in a position to put our ideas and findings to work.

Summaries, in html format, are available for many articles published since 2002.

www.newyorkfed.org/research/epr/index.html

FEDERAL RESERVE BANK OF NEW YORK
ECONOMIC POLICY REVIEW

Single-copy subscriptions to the *Economic Policy Review* (ISSN 1932-0426) are free. All articles can also be downloaded at no charge from the Research Group's website.

Library of Congress Card Number: 77-646559

Federal Reserve Bank of New York research publications—the *Economic Policy Review*, *Current Issues in Economics and Finance*, *Staff Reports*, *Research Update*, and the annual publications catalogue—are available at the Research Group's website.

www.newyorkfed.org/research