Federal Reserve Bank of Chicago

Call for Papers

2005 Conference on Bank Structure and Competition

Fourth Quarter 2004

Economic perspectives

2 Interest rates and the timing of new production

12 In search of a robust inflation forecast

34 How do banks make money? The fallacies of fee income

52 How do banks make money? A variety of business strategies

68 *Index for 2004*

Economic perspectives

Economic Perspectives is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Economic Perspectives and other Bank publications are available on the World Wide Web at http://www.chicagofed.org.

chicagofed.org

ISSN 0164-0682

Contents

4th Quarter 2004, Volume XXVIII, Issue 4

2 Interest rates and the timing of new production

Boyan Jovanovic and Peter L. Rousseau

This article studies the relation between IPO investment and the rate of interest. The 1950s and early 1960s, especially, were periods of very low real interest rates, and IPO investment was very low, with firms delaying their IPOs significantly. The authors find a qualitative difference between investment of IPO-ing firms and the investment of incumbent firms. The latter is decreasing in the interest rate, as neoclassical theory predicts. On the other hand, very low interest rates tend to discourage IPOs, and this may be why the 1950s and 1960s contained few IPOs.

12 In search of a robust inflation forecast

Scott Brave and Jonas D. M. Fisher

It is difficult to consistently improve upon forecasts of inflation based solely on the most recent data on inflation. In this article, we show how to do so. Our main finding is that the most robust forecasts combine information from several different forecasting models, each of which incorporates the information in the available inflation indicators in different ways.

32 *Conference on Bank Structure and Competition Announcement*

34 How do banks make money? The fallacies of fee income

Robert DeYoung and Tara Rice

In the first of two articles in this issue, the authors document the increasing importance of noninterest income at U.S. commercial banking companies and address two fundamental misunderstandings regarding this trend: the belief that fee-based activities provide more stable earnings than interest-based income and the belief that fee income flows chiefly from nontraditional, nonbanking activities.

52 How do banks make money? A variety of business strategies

Robert DeYoung and Tara Rice

In the second of two articles, the authors show that the business strategy chosen by a commercial banking company can have a substantial impact on its risk–return profile. Their analysis suggests that a wide variety of business strategies are likely to be financially viable in the future but, echoing concerns of others, they conclude that very small banking companies will face financial challenges regardless of the business strategy they select.

68 *Index for 2004*

Interest rates and the timing of new production

Boyan Jovanovic and Peter L. Rousseau

Introduction and summary

Policymakers are naturally interested in the effects of interest rates on various economic activities. This article studies how interest rates affect entrepreneurs' propensities to initiate new projects. Since the implementation of new ideas and production techniques is an important engine driving long-run economic growth, the effect of real rates on this activity should be of particular interest. This article illustrates that the effect of interest rates on the incentives to implement is not monotonic. Starting at high interest rates, a fall in the interest rate will spur entrepreneurs to implement projects more rapidly. But lowering interest rates even further will only persuade entrepreneurs to delay.

Ordinarily it would be difficult to measure the extent of delay, since we cannot easily identify when an economic agent first received the opportunity to bring a project to fruition. To get around this, we look at initial public offerings (IPOs). Although the decision to issue an IPO may reflect a host of considerations, Jain and Kini (1994) find that IPOs appear to be related to growth in investment and sales. More importantly, we can measure the amount of time that transpired between when a firm was founded or incorporated and when its IPO was issued, so we have a reasonable proxy for the delay time. Data on the time it takes firms to go public show a non-monotonic correlation between interest rates and the age at which the firm goes public. High rates of interest induce a delay and discourage investment for the usual reason, namely that when future income is discounted more heavily, it is not worthwhile to sacrifice current resources. Very *low* rates of interest, however, also discourage investment, because profits that are foregone during the delay are not as costly in comparison with the gains to delaying.

Chetty (2001) has shown that irreversibility of investment can lead to a non-monotonic relation between interest rates and investment. In his two-period model, if investment is postponed to the second period, the firm can better react to news about demand conditions. Aside from offering a different model, we also provide evidence on the non-monotonicity. In earlier work, Jovanovic and Rousseau (2001) show that the incentive to delay implementing a project gets stronger as the interest rate falls. In that paper, we also provide an information-theoretic rationale for the gains to waiting, but do not give any evidence.

The non-monotonicity of physical investment in the interest rate stems, ultimately, from the fact that the firm is giving up profits while it waits to implement its project. The decision to wait itself delivers information, that is, human capital, hence what is really happening is a substitution of one form of capital for another. We comment on this again in the conclusion and the implications that it may have for countries like Japan that are experiencing low investment, in spite of enjoying very low interest rates.

In the next section, we explain the model, and in the following section we describe its main implications for the data that we have. Then, we test those implications and discuss some related literature.

The model

The following model is a simplified version of Jovanovic and Rousseau (2001). Suppose the firm lives forever and has the property rights to its project. When implemented, the project produces output using

Boyan Jovanovic is a professor of economics at the University of Chicago and New York University, a research associate of the National Bureau of Economic Research (NBER), and a consultant to the Federal Reserve Bank of Chicago. Peter L. Rousseau is associate professor of economics at Vanderbilt University and a research associate of the NBER. *The authors thank the National Science Foundation for financial help.*

knowledge and physical capital k. The firm starts to receive net revenue cash only after it implements the project. Let T denote the waiting time until implementation. Suppose that while it waits, the firm's potential output is

$$y = f(T).$$

We assume that f increases with T but at a diminishing rate, as drawn in figure 1. In this formulation the firm starts receiving y only *after* implementing the project. At that point the project starts yielding profits. Moreover there are no direct costs. In that case the implementation decision is much like the decision of how long to remain in school. This is like perfecting an idea before taking out a patent on it.

Choosing the implementation date when there is no physical capital

If the firm lives forever and has the property rights to its project, it must just decide when to implement it. There are no direct costs. Only implicit "foregone-earnings" costs. The problem we analyze is similar to the well-known tree-cutting problem in economics, in which one wants to figure out the optimal time to cut down a tree. The trade-off involved is that between selling a young tree for cash today as opposed to selling a more mature tree for more cash tomorrow. The rate of interest has an important influence on that trade-off.

Formally, the firm's problem is that of choosing the implementation date T to maximize the present value of its future net revenues

$$e^{-rT}\frac{1}{r}f(T).$$

One can show that the optimal timing will satisfy the following equation:

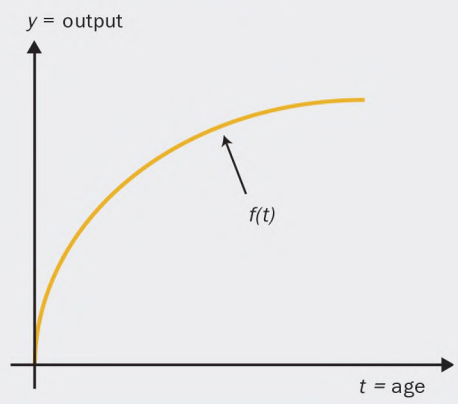
$$1)\quad f(T) = \frac{1}{r}f'(T).$$

The left-hand side of equation 1 is the foregone-earnings costs of waiting another period. In the problem as stated, this is the only cost. The right-hand side is the gain from waiting. Since this gain is received in every subsequent (production) period, it is capitalized, and hence the r in the denominator. It is more revealing to write the condition as

$$2)\quad g = r,$$

FIGURE 1

Post-IPO output of the firm as a function of its age at IPO



where

$$g \equiv \frac{f'(T)}{f(T)}$$

is the rate of growth of potential output. Thus, the implementation occurs when g equals the rate of interest.

Example

As an example, consider $f(t)=At^{\alpha}$, where $\alpha < 1$.

Here the condition reads $\frac{\alpha}{T} = r$, so that

$$3)\quad T = \frac{\alpha}{r}.$$

In this simple version of the model, then, a rise in the rate of interest hastens the implementation because it makes the foregone-earnings cost of waiting more important relative to the future gains from waiting. Interestingly, the productivity of the firm, A, does not affect the firm's implementation date because it simply scales both costs and revenues in the same proportion.

The parameter α will be important in what follows. It measures the gain in productivity that the firm gets by delaying its implementation. Delay lets the firm resolve technological uncertainties, perfect its ideas, and choose the right inputs for its production process.

Adding physical capital

To the extent that implementation entails spending on capital goods (as suggested by the evidence in Jain and Kini, 1994), this implies that the effect of the real rate of interest on investment is unambiguously

positive! Lower rates discourage implementation by inducing firms to wait longer so as to perfect their investments. The only cost is that of the profits that are postponed—a foregone-earnings cost.

In reality, firms must incur direct costs of implementation. However, these direct costs now introduce a new consideration: Higher interest rates imply it is better to defer these costs into the future since their present value is smaller. This suggests that lowering the interest rate will mitigate the incentive to delay, and that ignoring fixed costs of implementation (even if they do not correspond to measured investment) may be misleading. Therefore, we now introduce capital expenditure of I that is incurred at the implementation date. This modifies the firm's problem to one of choosing T to maximize the following present value:

$$e^{-rT}\left\{-I+\frac{1}{r}f(T)\right\}.$$

One can now show that the optimal timing will satisfy the following equation:

$$4)\quad rI-f(T)+\frac{1}{r}f'(T)=0,$$

so that instead of equation 2, the condition of optimality reads

$$5)\quad g=r-\left(\frac{I}{f(T)}\right)r^2.$$

Now g is essentially a quadratic in r. When r is small, the effect of r on g is positive as before, but when r gets large, the opposite is true, and the effect of r on g is non-monotonic. Note, too, that the coefficient on r^2 is the capital output ratio. As a result, the effect on T is non-monotonic too, and with it the effect on implementation investment.

The example again

To illustrate this, let us return to and augment the example $f(t) = At^{\alpha}$ we outlined above. The firm's problem becomes one of choosing T to maximize the following present value:

$$e^{-rT}\left\{-I+\frac{1}{r}AT^{\alpha}\right\}.$$

Figure 2 plots the optimal implementation delay on the vertical axis and the rate of interest on the horizontal axis. We see that for a smaller r, the term $2/r$ dominates, driving T to infinity. For a larger r, the term rI/A dominates, again driving T to infinity. We therefore have a U-shaped relation between r on the horizontal axis and T on the vertical, as illustrated in figure 2 for the case where $I = 30A$. We also plot T for the case where $I = 45A$, and $I = 60A$. We note that 1) the curves bottom out at levels of r ranging between 5 percent and 10 percent, and 2) higher investment outlays imply longer waiting at all levels of the interest rate. For practical purposes, however, the size of the outlay, I, starts to matter only when the interest rate is relatively high, say above 4 percent.

Implications of the model

The model has time-series and cross-sectional implications. The time-series implications concern low-frequency movements in T and the market value of the firm at IPO, which we denote as

$v=e^{-rT}\left\{-I+\frac{1}{r}AT^{\alpha}\right\}$. We are especially interested in the relation between interest rates and IPO investment. The model assumes that r is fixed, and therefore we may, at best, take figure 2 to predict the effects on T of low-frequency movements in r. These movements will induce changes in total investment spending—the total outlays on I—that we associate with implementation investment. The above framework lets us derive the following results.

FIGURE 2

Plot of T on r when $\alpha = ½$

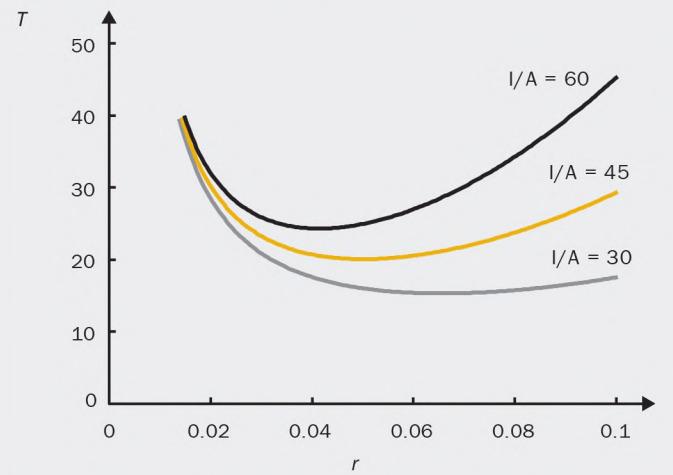
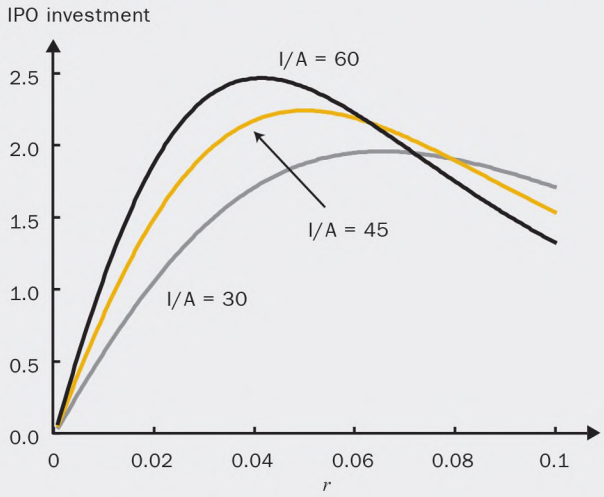


FIGURE 3

Backward bending investment schedules when $\alpha = ½$



Relationship between time to go public and the real interest rate

At low frequencies, the relation between T and r is U-shaped, as figure 2 shows. This means that the investment schedule is backward bending. We note that the negative relationship that emerges at low levels of the real rate is more pronounced than the positive relation at higher rates and that such high rates are not often observed.

Relationship between investment and the real interest rate

The results on the effects of r on T can now be translated into results for IPO investment. A rise in T means that investment is postponed. Consider the stock of new projects that need implementing. Into this stock there is an inflow of new projects as entrepreneurs get new ideas and at the same time an outflow due to projects being implemented. Investment will be proportional to the outflow of projects, because any project that is implemented requires investment. An increase in T will imply that the current cohort of projects will take a long time to leave. But if the inflow of ideas is constant, in the new steady state the outflow will be constant as well. Any effects of changes in T will only affect the transitional path.

The size of this transient effect will depend on the difference $T_{NEW} - T_{OLD}$. To see this more clearly, consider an economy that has a constant inflow of ideas. If a change in r (perceived by firms to be permanent) raises T, then strictly speaking we should see no investment at all for $T_{NEW} - T_{OLD}$ periods, followed immediately by the same steady state investment rate as took place before the change. Conversely, if a change in r lowers T, then there would immediately be a burst of investment that implements all existing ideas that are older than T_{NEW}. The general point is that interest-rate variation at low frequencies will produce changes in investment that are in the direction opposite to the change in T, and this change is related to the level of T_{NEW}.

Roughly speaking, then, decade to decade, we may expect a negative relation between T and implementation investment. Therefore, the relation between investment on the vertical axis and the rate of interest on the horizontal should have an inverted-U shape. We illustrate this in figure 3. The vertical axis shows the ratio I/T plotted against r by decade. The curves cross because T is increasing in I, and the ratios are not ordered the same way at different levels of r. But what is important here is the inverted-U shape in the graph and this is what we are looking for in the data.

IPO-issuing firms versus stock-market incumbents

Our model derives implementation lags from the improvement of projects prior to their implementation.

FIGURE 4

Waiting times to exchange listing, 1886–2002

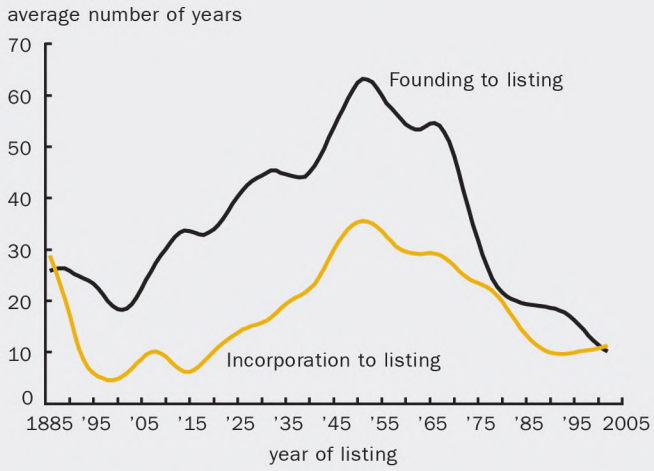


TABLE 1

Firms in the waiting-time sample

Decade	Number of new CRSP listings	Number of incorporations	Number of foundings
1890–99	112	52	41
1900–09	112	78	44
1910–19	214	190	97
1920–29	545	492	273
1930–39	231	197	78
1940–49	271	246	97
1950–59	254	241	78
1960–69	2,008	964	198
1970–79	4,517	1,405	262
1980–89	6,322	904	790
1990–99	7,850	1,539	1,939
2000–02	1,311	324	324
Total	23,747	6,632	4,221

It is the upward slope in figure 1 that creates the incentive for a firm to delay implementation while the project is improved and refined. The returns to waiting should, in turn, depend on how uncertain the environment is for the firm and its project. These uncertainties are likely to be greater for new products and new markets, and it is in such products and markets that new firms predominate. IPO-issuing firms tend to be new, or they at least tend to be younger than most established corporations. Therefore, we expect to see a difference between the investment behavior of entrants and incumbents.

The parameter that the model isolates in this regard is α. The curvature of f is likely to be larger, and the returns to waiting likely to be smaller, for established firms. This is most evident in equation 3, where a low α reduces the incentive to delay and therefore mitigates the forces that we have been describing here. In the expanded version of the model where we allow for physical investment, this simply means that the incentive to delay because of improving the project is weaker relative to the standard considerations of comparing I with discounted profits.

As a result, we expect to find a quantitative difference between the estimated investment schedules of incumbents and IPO-issuing firms. Even for incumbents, the incentives to delay should be there, but they should be much smaller. We thus expect to see less of a backward bend, if any, in the investment schedules of established firms.

Tests of the implications

Having listed the main implications of the model, we report on how they fare with the data, taking them up in the same order as above. IPOs provide a context for measuring a delay until investment—Jain and Kini (1994) find that IPOs are associated with a rise in investment and sales. Our use of IPO data in testing the theory is reasonable if:

1. Funds are a constraint for private companies;
2. IPOs can deliver the funds for a significant expansion; and
3. Upon the initial expansion, the firm is irrevocably defined and its IPO investments cannot be reversed.

When these assumptions hold at least approximately, we may interpret the firm's age at the IPO date as a proxy for the delay time to investment. Some of the costs incurred at IPO are transaction costs—Lee et al. (1996). We lump all costs into I and treat them as "investment."[1]

Testing the relationship between time to go public and the real interest rate

The first implication says that the relation between T and r should be U-shaped. To measure T, we construct average waiting times from founding and incorporation to stock-exchange listing since 1886,

FIGURE 5

The ex post real interest rate on commercial paper, 1885–2002

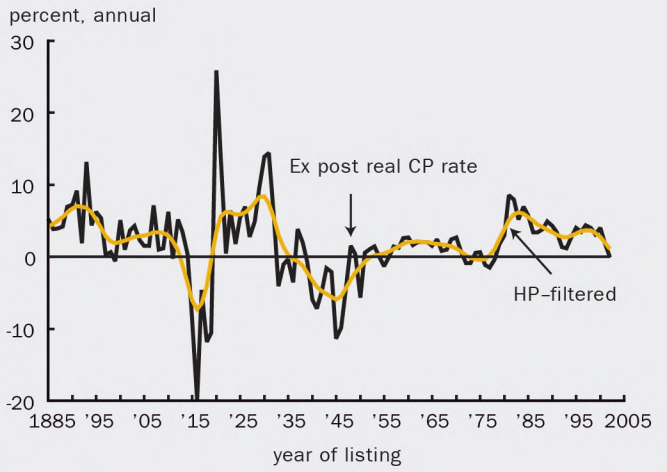
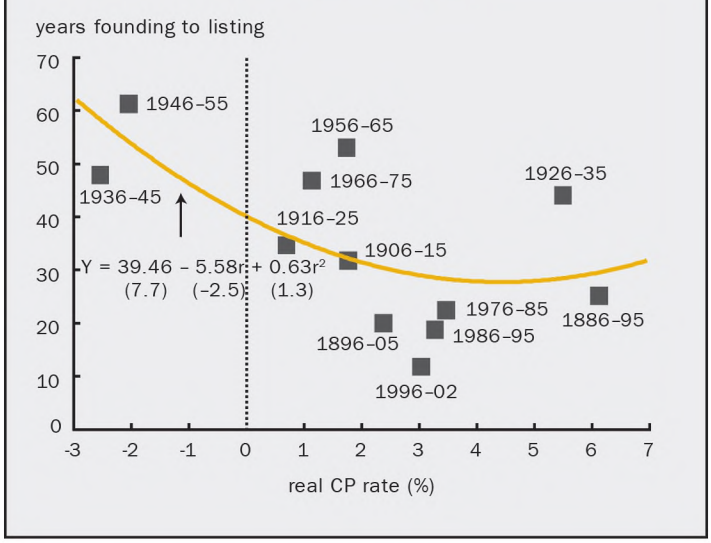


FIGURE 6

T (from founding) on *r*



based on individual company histories and our extension of the stock files distributed by the University of Chicago's Center for Research in Securities Prices (CRSP) from its 1925 starting date back through 1885 using newspaper sources.[2] Figure 4 shows these series after smoothing with the Hodrick-Prescott filter. Table 1 shows the coverage of our collection of IPO waiting times by decade. Waiting times by either measure were longest in the 1950s and 1960s and shortest at both ends of the twentieth century.

To what extent do these waiting times reflect waiting to implement projects? According to figure 4, the smoothed number of years between founding and listing ranges from ten to 60 years. It is hard to believe that a firm delays entirely for the purpose of perfecting and honing and then finally initiating its project when it goes public. Moreover, many profitable firms remain private. The time it takes to go public probably depends on several factors that are absent from our model. What matters, however, is time variation in the time to go public, which, barring any technological changes, is probably driven partly by incentives that we have modeled. While it may at first seem unlikely that the age at IPO should have increased by 15 years or 20 years in the 1940s entirely in response to interest rates, figure 2 shows that the model is able to generate very sharp increases in waiting times as interest rates near zero. Indeed, this is a robust implication. From equation 5 it follows that as the interest rate tends toward zero, the waiting time goes to infinity. No other parameter restrictions are required for this conclusion to hold. It is also true, however, that the relation is much steeper at low rates than it is at high rates. Thus, the greatest potential of this model to explain waiting times is when interest rates fluctuate around a low level.

Figure 5 shows the real interest rate on commercial paper with 30–90 days until maturity from 1885 to 2002, along with an HP-filtered (Hodrick–Prescott) trend.[3] Real rates were lowest in the middle of the twentieth century, and the series is roughly U-shaped. The long wait times in the 1950s and the corresponding negative real interest rates appear roughly consistent with our model. To examine the low-frequency relationship between *T* and *r* more precisely, however, we average both across ten-year periods and test for non-monotonicity with a quadratic regression.

Figure 6 shows a scatterplot of averages by decade of *T* on *r*, with *T* measured by the number of years from founding to exchange listing. Figure 7 instead uses years from incorporation as the measure of *T*. In either case, a U-shaped pattern appears in the data. The regressions in table 2 confirm this, with the coefficient on the real interest rate negative and significant at the 5 percent level for the linear term and positive (though

FIGURE 7

T (from incorporation) on *r*

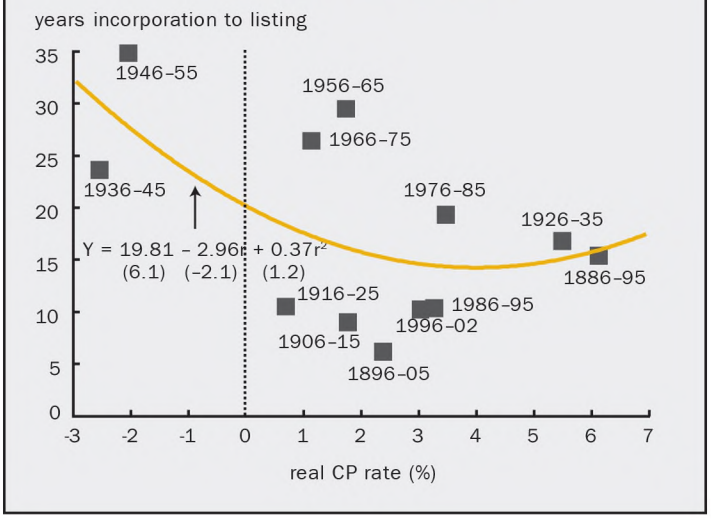


TABLE 2

Regressions of waiting times (T) on the real commercial paper rate (r) by decade, 1886–2002

	Dependent variable			
	T from founding		*T* from incorporation	
r_t	–3.47	–5.58	–1.71	–2.96
	(–2.23)	(–2.46)	(–1.77)	(–2.07)
r_t^2		0.63		0.37
		(1.25)		(1.17)
constant	41.68	39.46	21.11	19.81
	(8.37)	(7.65)	(6.80)	(6.10)
R^2	.33	.43	.24	.34
N	12	12	12	12

Note: T-statistics are in parentheses.

not significant) for the quadratic term. We interpret this as supporting evidence for the first implication of our model. We note, however, that negative real interest rates are inconsistent with the model and that instead of varying between 0 percent and 10 percent (as the interest rate does in the theoretical plots of figures 1–3), the decade averages vary from about –3 percent to 7 percent.

Testing the relation between investment and the real interest rate

The second implication deals with the relation between IPO investment and the real rate of interest. In testing for this, we provide a parallel analysis of the relation between aggregate investment (which is dominated by investment of stock-market incumbents) and the rate of interest. We do this to contrast the two relationships.

IPO-issuing firms probably face much greater uncertainty than incumbent firms. IPO-issuing firms are in the process of defining themselves, their products, and their technologies, and once they have chosen these directions, there is no going back for most of them. Choosing the wrong standard, for example, can condemn a new business to an early demise. There is a real sense, then, in which their investments are irreversible.

Incumbent firms, on the other hand, have chosen their domains of operation and face uncertainty more in the scale of demand, input prices, and so on. For these firms, there is less to be gained by waiting because there is less uncertainty to be resolved by delaying investment. Therefore, we would expect the investment of incumbents to be negatively related to the rate of interest. So, while we do not offer a model of incumbent investment, we note that the standard Q-theory model of investment (for example, Hayashi, 1982) with convex adjustment costs and no irreversibilities, predicts that a rise in the interest rate reduces investment.

Our model implies that, unlike incumbent investment, the relation between IPO investment and the rate of interest should be an inverted-U. Figure 8 shows the two investment series that we consider. The yellow line is private domestic investment as a percentage of the aggregate capital stock.[4] The black line is the value of IPO-issuing firms at the end of each year as a percentage of total stock market capitalization.[5] While investment rates tended to rise until the Great Depression and then stabilized after World War II, IPOs followed a more erratic pattern, with the value of new equity largest around the turn of the twentieth century, around 1915, in the late 1920s, at the end of World War II, in the late 1960s, the mid-1980s, and the 1990s.

To examine the low-frequency relationship between these measures of investment and r more precisely, we again average across ten-year periods.

Figure 9 shows a scatterplot of decade averages of r on IPO value, along with the fitted values from a quadratic regression. Figure 10 shows the scatterplot and quadratic regression line for incumbents' investments.

FIGURE 8

Annual IPOs as a % of stock market value; private investment as a % of the capital stock, 1886–2002

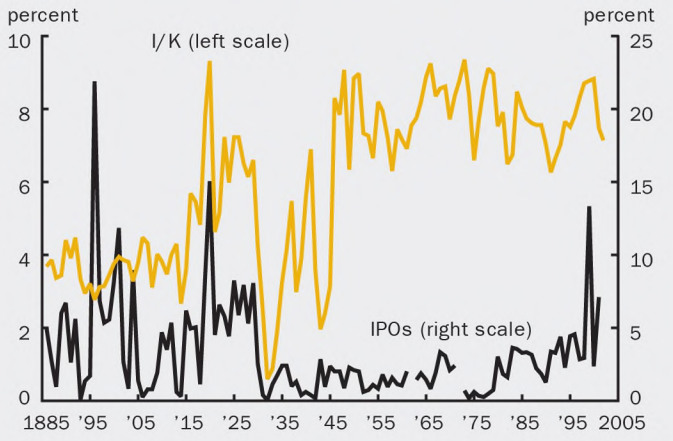
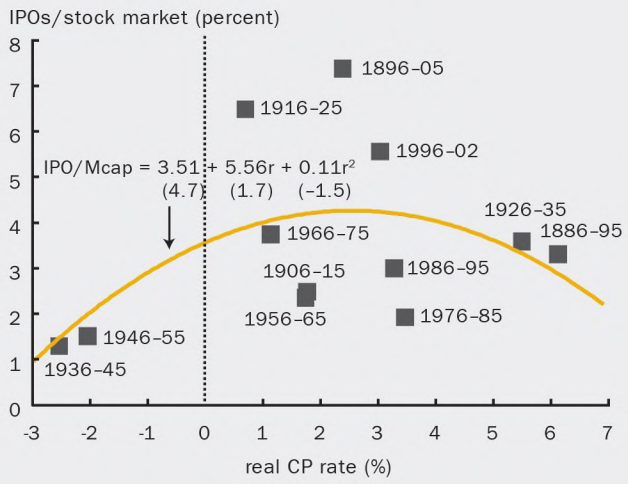


FIGURE 9

IPOs as a share of stock market capitalization



We report the details of the quadratic regressions and their linear counterparts in table 3. For IPO investment, the linear term is positive and statistically significant at the 5 percent level, while the coefficient on the quadratic term is negative and approaching statistical significance. We interpret this as evidence for the inverted U-shape that the model predicts. With incumbent investment, we also find an inverted U-shape, but the coefficient on the linear term is much smaller and not statistically significant.

Summary of the empirical results

To the extent that we may proxy implementation delays by the ages of firms at their IPOs, our results, on the whole, confirm the implications of the model. This is especially true for the backward-bending IPO-investment schedule. We did not find such evidence for the investment of established firms.

Our focus has been on the individual firm's decision and not the aggregate equilibrium aspects surrounding IPOs. Had we analyzed these, we would have needed to mention economies of scale in IPO activity and start-up activity (for example, due to concentration of venture capital focus) and to discuss the models of Diamond (1982) and Veldcamp (2003) that could perhaps explain some IPO waves.

We have assumed that, at IPO, the public pays exactly what the firm is worth. In a more expansive paper, one could entertain a hypothesis of "irrational exuberance," or times when the public is willing to pay more than the firm is worth. Along the lines of Shleifer and Vishny's (2003) paper on mergers, one could argue that perhaps IPO-issuing firms wait in the wings in order to take advantage of such exuberance. If so, the beneficiaries are neither the IPO-issuing firms nor the participating venture capitalists themselves. Data from Ritter (2003a, b) show that, despite being times of high IPO volume, high-*Q* periods are, in fact, times of more severe underpricing of firms going public. In other words, models in which a naive shareholder buys overpriced firms will not explain the time-series correlation between the volume of IPOs and Tobin's Q. Perhaps it is only the investment bankers who benefit from such exuberance.

Conclusion

We have presented and tested a neoclassical model with liquidity constraints. In this model, delay to implementation occurs because the firm is trying to improve its idea to the point where it becomes optimal to incur the fixed cost of implementing a project.

The broader implication of our work here is that lowering interest rates may impede new ideas rather than foster them. But this does not mean that low interest rates are bad for firms, even when they lead firms to postpone their investment. Regardless of

FIGURE 10

Investment rate on *r*

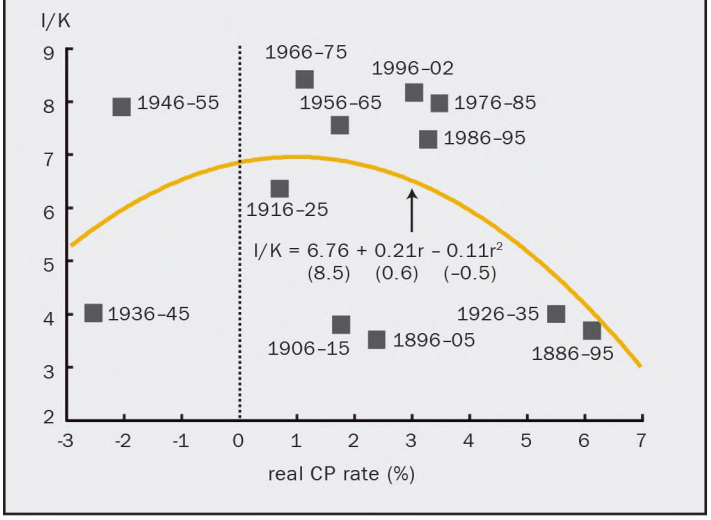


TABLE 3

Regressions of IPOs and the investment rate on the commercial paper rate (r) by decade, 1886–2002

	Dependent variable			
	IPOs / Stock market		I/K	
r_t	0.20	0.56	–0.16	0.21
	(0.86)	(1.71)	(–0.66)	(0.61)
r_t^2		–0.11		–0.11
		(–1.48)		(–1.45)
constant	3.13	3.51	6.37	6.76
	(4.23)	(4.71)	(8.08)	(8.49)
R^2	.07	.25	.04	.22
N	12	12	12	12

Note: T-statistics are in parentheses.

how investment reacts, the value of projects rises as the interest rate falls.

Nor do our results say that low interest rates discourage all investment broadly defined. Our finding that at low rates *physical*-capital investment rises with the interest rate is really about the composition of capital. A delay is a switch of one kind of investment profile for another. When the reason for delaying is the gathering of information, total investment (including information investment) may still be monotone-decreasing in the interest rate. Firms postpone physical investment, but they gather information, and this is human capital. Before implementing its project, the value of that project is monotone-decreasing in the interest rate, and that value—that is, the value of the physical and human capital combined—is being maximized by the firm's policy. Thus, when physical investment rises with the interest rate, this simply means that the firm's human capital investment is falling, and perhaps its total capital properly measured. Therefore, for example, the Japanese economy may be in better shape than it seems today because the very individuals that are not investing may be accumulating a different kind of capital that is not measured as such.

NOTES

[1]Other evidence shows that increasing funds for investment is indeed one of the motives behind an IPO. Jain and Kini (1994, table 2), for example, find that by the fourth year after its IPO, the firm will experience a rise in sales of 80 percent compared with its industry counterparts and 143 percent compared with its own sales in the year just before the IPO (also see Choe, Masulis, and Nanda, 1993; Lowry, 2002; and Moskowitz and Vissing-Jorgensen, 2002). We find that the 1955–2001 correlation between funds that firms take in at IPO and their real investment is 0.33 and highly significant.

Our assumption that the firm's investment occurs at the time of IPO brings us closer to the literature on liquidity constraints. When an entrepreneur has a high return activity that he cannot fund in the capital market, he has a greater incentive to save, because those savings can fund an investment that is more profitable than the average market investment. Buera (2003) analyzes optimal saving behavior by liquidity-constrained entrepreneurs.

[2]Listing years after 1925 are those for which firms enter CRSP. For 1885–1924, they are years in which prices first appear in the New York Stock Exchange (NYSE) listings of *The Annalist*, *Bradstreet's*, *The Commercial and Financial Chronicle*, or *The New York Times*. The 6,632 incorporation dates used to construct figure 4 are from *Moody's Industrial Manual* (1920, 1928, 1955, 1980), Standard and Poor's *Stock Market Encyclopedia* (1981, 1988, 2000), various editions of *Standard and Poor's Stock Reports*, and *Mergent Online*. The 4,221 foundings are from Dun and Bradstreet's *Million Dollar Directory* (2000), Moody's, Etna M. Kelley (1954), and individual company websites. We linearly interpolate the series between missing points before applying the HP-filter to create the time series in the figure.

[3]Commercial paper rates are annual averages of 30-day terms from the FRED (Federal Reserve Economic Data) database for 1934–2002 and 60–90 day terms from Homer and Sylla (1991) for earlier years. We compute the ex post return by subtracting inflation as computed by the growth of the implicit price deflator for gross domestic product (GDP) from the U.S. Bureau of Economic Analysis (BEA) (2003) for 1929–2002 and Berry (1988) for earlier years.

[4]To build the investment rate series, we start with gross private domestic investment in current dollars from the U.S. Bureau of Economic Analysis (2003, table 1, pp. 123–124) for 1929–2001 and then ratio-splice the gross capital formation series in current dollars, excluding military expenditures, from Kuznets (1961b, tables T-8 and T-8a) for 1870–1929. We construct the net capital stock using the private fixed assets tables of the Bureau of Economic Analysis (2003) for 1925–2002. Then, using the estimates of the net stock of non-military capital from Kuznets (1961a, table 3, pp. 64–65) in 1869, 1879, 1889, 1909, 1919, and 1929 as benchmarks, we use the percent changes in a synthetic series for the capital stock formed by starting with the 1869 Kuznets (1961a) estimate of $27 billion and adding net capital formation in each year through 1929 from Kuznets (1961b) to create an annual series that runs through the benchmark points. Finally, we ratio-splice the resulting series for 1870–1925 to the later BEA series. The investment rate that appears in figure 8 is the ratio of our final investment to the capital stock series, expressed as a percentage.

[5]The stock market data are from the CRSP files and our backward extension of them to 1885. NYSE firms are available in CRSP continuously, AMEX firms after 1961, and NASDAQ firms after 1971. New listings are given by the total year-end market value of firms that entered our database in each year, excluding American depository receipts (ADRs).

REFERENCES

Berry, T. S., 1988, *Production and population since 1789: Revised GNP series in constant dollars*, Richmond, VA: The Bostwick Press, Bostwick Paper, No. 6.

Bradstreet Co., 1885–1925, *Bradstreet's*, New York: Bradstreet Co., various issues.

Buera, Francisco, 2003, "A dynamic model of entrepreneurship with borrowing constraints," University of Chicago, manuscript, November.

Chetty, Raj, 2001, "Interest rates, learning, and business investment under uncertainty," University of California, Berkeley, manuscript, December.

Choe, Hyuk, Ronald Masulis, and Vikram Nanda, 1993, "Common stock offerings across the business cycle," *Journal of Empirical Finance*, Vol. 1, pp. 3–31.

Diamond, Peter, 1982, "Aggregate demand management in search equilibrium," *Journal of Political Economy*, Vol. 90, No. 5, October, pp. 881–894.

Dun and Bradstreet, Inc., 2003, *D&B Million Dollar Directory*, Bethlehem, PA: Dun and Bradstreet Inc.

Federal Reserve Bank of St. Louis, 2003, "Federal Reserve economic data (FRED)," database.

Hayashi, Fumio, 1982, "Tobin's marginal q and average q: A neoclassical interpretation," *Econometrica*, Vol. 50, No. 1, January, pp. pp. 213–224.

Homer, Sidney, and Richard Sylla, 1991, *A History of Interest Rates*, third edition, New Brunswick, NJ: Rutgers University Press.

Jain, Bharat A., and Omesh Kini, 1994, "The post-issue operating performance of IPO firms," *Journal of Finance*, Vol. 49, December, pp. 1699–1726.

Jovanovic, Boyan, and Peter L. Rousseau, 2001, "Why wait? A century of life before IPO," *American Economic Review*, Papers and Proceedings, Vol. 91, No. 2, May, pp. 336–341.

Kelley, E. M., 1954, *The Business Founding Date Directory*, Scarsdale, NY: Morgan and Morgan.

Kuznets, Simon S., 1961a, *Capital in the American Economy: Its Formation and Financing*, Princeton, NJ: Princeton University Press.

_____, 1961b, "Annual estimates, 1869–1955," Johns Hopkins University, Baltimore, MD, manuscript.

Lee, Inmoo, Scott Lockhead, Jay Ritter, and Quanshui Zhao, 1996, "The costs of raising capital," *Journal of Financial Research*, Vol. 19, No. 1, Spring, pp. 59–74.

Lowry, Michelle, 2002, "Why does IPO volume fluctuate so much?," *Journal of Financial Economics*, Vol. 67, No. 1, pp. 3–40.

Moody's Investors Service, 1920, 1929, 1931, 1941, 1951, 1956, 1961, *Moody's Industrial Manual*, New York: Moody's Investors Service.

Moskowitz, Tobias J., and Annette Vissing-Jorgensen, 2002, "The returns to entrepreneurial investment: A private equity premium puzzle?," *American Economic Review*, Vol. 92, No. 4, September, pp. 745–778.

National News Service, 1885–1925, *The Commercial and Financial Chronicle*, New York, various issues.

New York Times Co., The, 1897–1928, *The New York Times*, New York, various issues.

_____, 1913–1925, *The Annalist: A Magazine of Finance, Commerce, and Economics*, New York, various issues.

Ritter, Jay, 2003a, "Long-run returns on IPOs from 1970–2002," University of Florida, manuscript.

_____, 2003b, "Some factoids about the 2002 IPO market," University of Florida, manuscript, January.

Shleifer, A., and R. W. Vishny, 2003, "Stock market driven acquisitions," *Journal of Financial Economics*, Vol. 70, No. 3, December, pp. 295–311.

Standard and Poor's Corporation, 1981, 1988, 2000, 2002, *Stock Market Encyclopedia*, New York.

_____, various years, *Stock Reports*, New York: Standard and Poor's Corporation.

United States Bureau of Economic Analysis, 2003, *Survey of Current Business*, Washington, DC: Government Printing Office.

University of Chicago, Center for Research on Securities Prices, 2002, *CRSP Database*, Chicago.

Veldcamp, Laura, 2003, "Media frenzies in markets for financial innovation," New York University, manuscript, June.

In search of a robust inflation forecast

Scott Brave and Jonas D. M. Fisher

Introduction and summary

The sound conduct of monetary policy is the bedrock on which a well-functioning economy rests. In the United States, the conduct of monetary policy is guided by the goals set out in the 1977 amendment to the Federal Reserve Act of 1913. According to this amendment, the Federal Reserve System and the Federal Open Market Committee (FOMC) should conduct monetary policy to promote the goals of "maximum" employment and output and to promote "stable" prices.

Of these goals, the primary focus, many economists believe, should be on achieving price stability. A stable price level means that prices of goods and services are undistorted by inflationary surprises. This enhances the role of prices in providing signals to ensure the efficient allocation of resources and the maximum possible sustainable level of employment. Many also believe that a stable price level encourages saving and capital accumulation, because it prevents asset values from being eroded by unanticipated inflation or debt being amplified by unanticipated deflation. This should also contribute to the goals of attaining maximum employment and output.

For these reasons, monetary policy is heavily influenced by factors thought to affect the rate of change of prices, that is, inflation. Until recently, the dominant concern had been a recurrence of past episodes of high inflation that have been associated with bad macroeconomic outcomes. In recent years, however, concern has shifted to the possibility of deflation. In either case, given the long lags over which policy actions can take effect, it is often necessary for the FOMC to take action before inflation starts to move in an undesired direction. The only way to do this with some confidence is to have effective ways of predicting the future course of inflation. Hence, forecasting inflation is a crucial ingredient in the formulation of monetary policy.

This article is concerned with the ability to forecast inflation. This is a relevant issue since recent work has cast doubt on the reliability of traditional approaches to forecasting inflation. Inflation forecasting is usually conducted with statistical models based on some version of the Phillips curve, the statistical relationship between inflation and overall aggregate economic activity. The recent literature suggests that this approach has not been reliable. In particular, Atkeson and Ohanian (2001) found that over the period 1985–99, one-year-ahead forecasts of inflation based on the Phillips curve do no better than a "naive" forecast where the forecast is set to the inflation rate over the prior year.

Some researchers have come to the defense of traditional forecasting models, arguing that the failure pointed out by Atkeson and Ohanian (2001) is special to the sample period they consider.[1] Still, it is difficult to dismiss their finding out of hand. As is clear from the work of Stock and Watson (1999, 2002, 2003), the forecasting failure in the post-1985 period reflects a more fundamental problem. While particular inflation forecasting models may do well in some periods, more often than not these models perform poorly at other times. It is not enough for a forecasting model to do well in just the recent period, because it is also important to guard against the possibility of structural change. Forecasters need to know that their forecasting strategy is robust to changes in the economic environment that are not noticed until well after they have occurred.

This article, therefore, addresses the question: Is it possible to build a robust inflation forecasting framework that does well in the recent period as well

Scott Brave is an associate economist and Jonas D. M. Fisher is an economic advisor at the Federal Reserve Bank of Chicago. *The authors thank Craig Furfine and Marcelo Veracierto for helpful comments.*

as earlier periods? We find that the answer to our question is "yes," although the gains compared with models based only on past inflation are at times quite modest. However, around periods in which inflation begins to pick up, the best models we consider show clear advantages over inflation-only models.

We address our question by considering the out-of-sample forecasting performance of a large set of models. We study forecast errors for the one-year and two-year forecasting horizons and at the monthly and quarterly frequencies. Our notion of robustness is that the model consistently lies near the top of performance lists of alternative models and is consistently more successful than models based only on past inflation, such as Atkeson and Ohanian's naive model.

Our main findings are as follows. First, consistent with previous studies, we show that different inflation indicators do well at forecasting inflation at different times. This makes the basic point that one should not rely on the "indicator du jour" when assessing the inflation outlook and that forecasters should be looking at many different indicators.

Second, we show that individual forecasting models that combine data in different ways do not consistently outperform the naive model (which turns out to be superior to other inflation-only models) in terms of mean-squared errors. For example, in some periods the naive model is better; at other times there is at least one model that does better than the naive model, but it is never the same one. This is true at both the one-year and the two-year horizon and with monthly and quarterly data. These findings are consistent with those reported by Fisher, Liu, and Zhou (2002).

Third, we show that certain kinds of models based on weighted averages of forecasts from individual models consistently outperform the naive model and other models based only on past inflation. This is true for both monthly and quarterly data and at both forecast horizons. At the one-year horizon, the best model involves weights computed using the within-sample forecasting performance of the individual models. At the two-year horizon, the best model uses a simple average of the individual models. For both forecasting horizons, the best versions of these models use a rolling window of data for the forecast, and these models are typically superior to the individual models for all sub-samples considered. These findings lead us to conclude that the most robust forecasts combine information from several different forecasting models, each of which incorporates the information in the available inflation indicators in different ways.

Another finding is that data available at the quarterly frequency that are not available at the monthly frequency appear to add little additional information to our forecasts. This might seem surprising, given that existing theoretical models suggest that data on real unit labor costs and productivity should be useful for predicting inflation, and these data are only available at the quarterly frequency. Still, we find that the additional data do not improve our forecasts very much, suggesting that most of the information about future inflation in the quarterly data is already incorporated in the monthly series we consider.

Below, we describe the different models we consider. Then, we discuss the methodology for assessing the forecasting performance of these models and present our findings.

Statistical models of inflation

In order to leave no stone unturned in our quest for a robust framework for forecasting inflation, we consider a large number of models. These models involve different ways of incorporating the vast amount of data available to the inflation forecaster. In principle, almost all the available macroeconomic data contain *some* information about future inflation. The challenge is to find a way to incorporate this information into a forecasting model. There are many ways to do this. One way would be to summarize the information useful for forecasting inflation *before* it is put into a model. Another approach would be to summarize the relevant information *after* it has been included in individual models. We employ each of these methods and also combine aspects of both. Finally, we combine the forecasts from several different types of models, each of which involves a different approach to forecasting. In the sub-sections that follow, we describe examples of each of these approaches. Many of these examples are motivated by the work of Stock and Watson (1999, 2002, 2003). For convenience we focus on the monthly frequency case. It should be clear how to extend the models to the quarterly frequency case. Table 1 summarizes the models underlying our analysis.

The basic regression equation

All the models we consider have as their foundation the basic regression equation:

$$1)\quad \pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{K} \theta_i(L) x_{it} + \varepsilon_{t+J},$$
$$J = 12, 24.$$

This equation relates changes in the 12-month inflation rate, defined as the 12-month change in the natural logarithm of the price index p_t,

$$\pi^{12}_{t} = \ln p_t - \ln p_{t-12},$$

TABLE 1

Summary of models

Model	Estimation equation	Indicators used
Naive	$\pi^{12}_{t+J} - \pi^{12}_{t} = \varepsilon_{t+J}$	None
Autoregression	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \varepsilon_{t+J}$	None
Natural rate	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Filtered unemployment rate
Output gap	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Filtered real GDP
Activity	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Index based on indicators listed in appendix
Indicator	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{3}\theta_i(L)x_{it} + \varepsilon_{t+J}$	Change in fed funds rate, unemployment rate, indicators listed in appendix
Combination	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Indicators listed in appendix
Diffusion	$\pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{K}\theta_i(L)x_{it} + \varepsilon_{t+J},\ K = 1, 2, \ldots, 6$	Six indexes based on indicators listed in appendix

Notes: See the text for a description of the notation and terminology. NA denotes not applicable; GDP denotes gross domestic product.

to past values of the one-month inflation rate, π_t,

$$\pi_t = \ln p_t - \ln_{pt-1},$$

and past values of other variables deemed useful for forecasting inflation, x_{it}, $i = 1, 2, \ldots, K$. In equation 1, α is a constant and $\beta(L)$ and $\theta_i(L)$, $i = 1, 2, \ldots, K$, specify the number of lags in inflation and other variables included in the equation. The number of other variables included is given by K, which is greater than or equal to zero.[2] We estimate equation 1 by ordinary least squares and use a standard lag selection criteria to choose the number of lags of inflation and other variables.[3] We allow for the possibility that lags could vary from one month to a year.

For given estimates of the coefficients in equation 1 at date $T, \hat{\alpha}_T, \hat{\beta}_T(L)$, and $\hat{\theta}_{iT}(L)$, the date T forecast of 12-month inflation J periods ahead using the basic regression equation is[4]

$$2)\quad \hat{\pi}^{12}_{T+J} = \pi^{12}_{T} + \hat{\alpha}_T + \hat{\beta}_T(L)(\pi_T - \pi_{T-1}) + \sum_{i=1}^{K}\hat{\theta}_{iT}(L)x_{iT},$$
$$J = 12, 24.$$

Models based only on inflation

We consider two models based only on inflation. The first is the "naive" model described by Atkeson and Ohanian (2001). The naive model can be viewed as a special case of equation 1, where $\alpha_T = \beta_T(L) = K = 0$. That is, the naive model equates the date T forecast of inflation over the next 12 months, $\hat{\pi}^{12}_{T+12}$, with its value over the most recent 12-month period,

$$3)\quad \hat{\pi}^{12}_{T+12} = \pi^{12}_{T}.$$

Similar to the 12-month forecast, the naive model equates the date T forecast of 12-month inflation 24 months into the future, $\hat{\pi}^{12}_{t+24}$, with its most recent value:

$$4)\quad \hat{\pi}^{12}_{t+24} = \pi^{12}_{t}.$$

The other model based only on inflation is called the *autoregression model*. This model postulates that changes in 12-month inflation only depend on recent changes in one-month inflation, that is, it sets $K = 0$ in equation 1.

Single equation models with inflation indicators

We consider three models that involve implementing equation 1 with $K = 1$. For the *natural rate model*, x_{1t} is set equal to the difference between a measure of the actual unemployment rate and an estimate of the "natural rate."[5] The *output-gap model*, is similar. In particular, x_{1t} is set equal to the difference between a measure of aggregate output and an estimate of "potential" output, where the latter is estimated using the same approach as with the natural rate.

For the *activity model*, x_{1t} is the Chicago Fed National Activity Index (CFNAI). This index is a weighted average of 85 monthly indicators of real economic activity. The CFNAI provides a single, summary measure of a common factor in these national economic data. As such, historical movements in the CFNAI closely track periods of economic expansion and contraction.[6]

Multiple equation models with inflation indicators

We also consider models that combine forecasts from applying versions of equation 1 with different indicator variables. The *diffusion model* can be viewed as a generalization of the activity model. We use a small number of indexes that explain the movements in 145 macroeconomic time series, including data measuring production, labor market status, the strength of the household sector, inventories, sales, orders, financial markets, money supply, and price data. The procedure that obtains the indexes processes the information in the 145 series, so that each index is a weighted average of the series and each index is statistically independent of the others. We consider six indexes computed in this way, d_{1t}, d_{2t}, ..., d_{6t}. These are listed in descending order in terms of the amount of information embedded in them.[7] The diffusion model involves first calculating an inflation forecast based upon including x_{1t} equal to the index with the most information, d_{1t}. We repeat this exercise five times, successively including one more index in descending order of importance. For instance, the third forecast created includes the three most important indexes, d_{1t}, d_{2t}, and d_{3t}, as x_{1t}, x_{2t}, and x_{3t}. The forecast from the diffusion model is the median of these six forecasts.[8]

Consider a list of forecasts of 12-month inflation J periods ahead at date T. Index these forecasts by n and denote them $f_{T+J}(n)$. The *combination model* is the median of these forecasts,

$$5)\quad \hat{\pi}^{12}_{T+J} = \text{median}\left\{ f_{T+J}(n) : n \in S \right\},$$

where the set of forecasts, S, is derived from the same 145 variables used to compute the diffusion indexes. In particular, each forecast $f_{T+J}(n)$ is based on equation 1 with $K = 1$ and x_{1t} set equal to one of the 145 variables used in the diffusion model.

The *indicator model* is based on a smaller list of variables grouped into six categories: economic activity, slackness measures, housing and building activity, industrial prices, financial markets, and, for the quarterly case only, productivity and marginal cost. Within each group, we compute a forecast using equation 1 with $K = 3$, x_{1t} set equal to the change in the federal funds interest rate, x_{2t} set equal to the unemployment rate, and x_{3t} to one of the variables in the group of indicators. We average the forecasts within each group. Then the indicator model forecast is based on equation 5 with $f_{T+J}(n)$ corresponding to one of the average forecasts from the five categories and S corresponding to the set of five average forecasts.

The combination and indicator models are useful to consider since they represent two alternatives to index-based methods for summarizing the information in many variables. The combination model is directly comparable to the diffusion model in that it involves the same set of variables. Therefore, it is useful to assess which method is superior for incorporating the information in a large number of variables. We work with the indicator model for two reasons. First, experience has shown it to be a relatively reliable approach to forecasting. Second, since it involves a small list of indicators, it represents a compromise between models that put a lot of weight on a single indicator, such as the natural rate and output gap models, and models that take virtually no stand on which indicators are useful, such as the diffusion and combination models.

Meta models

The preceding discussion introduced six models in addition to the inflation-only naive and autoregression models. To summarize, these models are the natural rate, output gap, activity, diffusion, combination, and indicator models. As we show below, none of these models consistently outperforms the inflation-only models over the various sub-samples we consider. However, for most of the sub-samples, at least one of the models does outperform the inflation-only models. This raises the question of whether it is possible to combine the information in these individual models to arrive at a superior forecast. The final group of models we study are designed to do just this. We call them *meta models*.[9]

Consider a list of forecasts of 12-month inflation J periods ahead at date T generated by the models listed above. Index these forecasts by n and denote them $f_{T+J}(n)$. The forecast of a given meta model is

$$6)\quad \hat{\pi}^{12}_{T+J} = \sum_{n \in M} w_{n,T} f_{T+J}(n),$$

where M is the set of models from which the meta model is constructed and $w_{n,T}$ is the weight attached to model n at date T. Equation 6 says that the forecast is set equal to a weighted average of the forecasts of the models comprising the meta model.

The meta models we consider differ according to the set of models from which the forecast is constructed and the manner in which the weights are computed. In the *equally weighted* models, the weights are all set equal to the inverse of the number of models comprising the model. That is, these forecasts are just the average over the forecasts of the individual models. The *optimally weighted* meta models have weights computed for each forecast date. These weights are computed as follows. At each forecast date, there is a prior history of forecasts and a history of actual inflation realizations corresponding to these forecasts. We reset the weights in equation 6 each forecast date to equal the coefficients of a regression of realized inflation on the forecasts using data on these variables available up to the date of the forecast.

Model evaluation methodology

We evaluate the accuracy of the models by comparing them with the naive and autoregression models. A modeling strategy will be deemed to be "robust" if it lies near the top of performance rankings and outperforms models based only on past inflation consistently across the various sub-samples we consider. We assess performance by *simulated out-of-sample forecasting*. This involves constructing inflation forecasts that a model would have produced had it been used historically to generate forecasts of inflation. We study forecasts of personal consumption deflator inflation, excluding food and energy, that is, core personal consumption deflator inflation.[10]

Two drawbacks of this approach are 1) we assume all the data are available up to the forecasting date, and 2) we do not use real-time data in our forecasts.[11] On a given date particular data series may not yet be published. Also many data series are revised after the initial release date. In our forecasting exercises, we compute forecasts and calculate the CFNAI and diffusion indexes assuming all the series underlying the forecasts and the indexes are available up to the forecast date. In practice this is never the case. Since we do not use real-time data, we also abstract from problems associated with data revisions. We suspect 1) and 2) lead us to overstate the effectiveness of our models.[12]

Root mean-squared error criterion

Our performance measure is the standard *root mean-squared error* (RMSE) criterion. The RMSE for any forecast is the square root of the mean squared differences between the actual inflation rate and the predicted inflation rate over the period for which simulated forecasts are constructed. For $J = 12, 24$

$$7)\quad RMSE = \left(\frac{1}{T-J} \sum_{t=1}^{T-J} \left[\pi^{12}_{t+J} - \hat{\pi}^{12}_{t+J} \right]^2 \right)^{1/2},$$

where $T-J$ denotes the number of forecasts made over the period under consideration.[13]

An advantage of the RMSE measure of performance is that its units are the same as inflation. This means, for example, the magnitude of RMSE for a given model can be directly compared with the average rate of inflation over the sample period. Another advantage is that large forecast errors are given more weight than small errors. Presumably, we care more about large mistakes than small mistakes. At the same time, a potential drawback of the RMSE measure is that it weights positive and negative errors of the same size in the same way. If we are more concerned about inflation increases than decreases, then this is definitely a drawback. Recent debates about the possible perils of deflation suggest that inflation decreases, at least at low levels of inflation, are certainly a concern of policymakers and so they should not be ignored. It would be interesting to consider other measures of forecast performance that weight increases and decreases in inflation differently, depending on the prevailing level of inflation.

Data and sample periods

The data we use in the analysis are described in the data appendix. The sample period of our analysis begins in 1967. We choose this date because it is the beginning date for the data used to construct the CFNAI and the diffusion indexes. We estimate the forecasting equations using all the data available at the time of the forecast and also consider the method of *rolling regressions*. A rolling regression keeps the number of observations in the regression constant across forecasts. Since it excludes observations from the distant past, this approach can in principle accommodate the possibility that there has been structural change in the data-generating

TABLE 2

Top five indicators, various sample periods: Combination and indicator variables

A. One-year ahead forecasts	B. Two-year ahead forecasts
1977–84	**1977–84**
ISM: Mfg: Prices Index	ISM Mfg: PMI Composite Index
Real inventories: Mfg: Durable goods industries	ISM: Mfg: Supplier Delivery Index
Housing starts: Northeast	ISM: Mfg: Inventories Index
ISM: Mfg: Inventories Index	ISM: Mfg: Employment Index
ISM: Mfg: Supplier Delivery Index	Housing starts: Midwest
1985–92	**1985–92**
Housing starts: Midwest	Housing starts: Midwest
NBER XLI2	Civilians unemployed for 15–26 weeks
Gold prices	Gold prices
Silver prices	Silver prices
CRB Futures Index	New home sales
1993–2000	**1993–2000**
Civilians unemployed for 5–14 weeks	Civilians unemployed for 5–14 weeks
Housing starts	Housing starts
3-year/1-year T-bill spread	Civilians unemployed for 15–26 weeks
10-Year Treasury note yield – federal funds rate	Housing starts: South
Civilians unemployed for 15–26 weeks	Building permits
2001–03	**2001–03**
Civilians unemployed for 27 weeks and over	Civilians unemployed for 5–14 weeks
Average duration of unemployment	Civilian unemployment rate: 16yr+
Civilians unemployed for 15 weeks and over	Employment retail and wholesale trade
Civilians unemployed for 5–14 weeks	Industrial Production Index
10-Year Treasury note yield – federal funds rate	Civilians unemployed for 15–26 weeks

process. To implement the rolling regression procedure, we choose a sample length of 15 years.

Finally, we consider four distinct periods over which to evaluate the forecasts of the models: 1977–84, 1985–92, 1993–2000, and 2001–2003. The first three periods are all 96 months long. We also consider the 1985–2003 period. The 1977–84 period is a period of high inflation volatility and general economic turbulence. The 1985–92 period is generally associated with a new monetary policy regime. This period also includes a mild recession. The 1993–2000 period witnessed uninterrupted economic expansion, stable monetary policy, and declining inflation. The 2001–2003 period is interesting because it involves recent forecast performance.

Findings

Next, we describe our findings. We focus on the monthly results and only discuss the findings with quarterly data at the end.

The best indicator keeps changing

Before evaluating our models, it is useful to consider the forecast performance of individual indicators. Each forecast is based on equation 1 with $K = 1$ and x_{1t} set equal to one of the list of indicators that includes the union of the set of variables used in the indicators model and the combination (or diffusion) model. Table 2 shows the top five indicators for the sample periods 1977–84, 1985–92, 1993–2000, and 2001–03. The key thing to notice from this table is that the list keeps changing! In the earliest sub-sample, indicators of manufacturing activity seem to do best at both the one-year and two-year horizons. At other times, employment, housing, or financial indicators do well. Overall, variables that do well at the one-year horizon do not necessarily do well at the two-year horizon. The lesson to be learned here is: beware of the indicator du jour.[14]

The best model keeps changing, too

Table 3 (p. 19) shows the performance of all the models (except for the output-gap model, which we only consider at the quarterly frequency) for the one-year and two-year forecast horizons, respectively. The meta models are in bold type. We discuss these models in the following sub-section. In table 3, we list the models for the four sub-samples as well as the period 1985–2003. We also display some useful summary statistics. For each sample period, we show the RMSE of

the best model, the range of RMSE across forecasting models, the absolute value of the difference between the naive model and the best model, and average actual inflation.

The first thing to notice is that for both forecast horizons and across all sample periods the naive model performs better than the autoregression model. That is, there is no more information about future inflation in past inflation than that already contained in the most recent reading of 12-month inflation. This fact motivates our focus on using the naive model as a benchmark for comparison.

Now, consider the one-year ahead forecasts. In the earliest period, 1977–84, the natural rate model performed best. The magnitudes of the errors from this forecast are about one-sixth of the average inflation rate in this period. This is large relative to the amount by which this best model outperforms the naive model; the difference between the best model and the naive model is only about one-thirtieth of the average inflation rate in this period. So, even in this early period, the naive model is difficult to beat.

Since 1985, it has been even harder to beat the naive model. Indeed, over the entire 1985–2003 period the naive model is the best performer of the individual models. Consistent with the findings in Fisher, Liu, and Zhou (2002), the success of the naive model is concentrated in the 1985–92 period. In the latter part of the post-1985 sample, there is a model that beats the naive model, but this model changes and the extent of the victory is quite small. We should not attribute too much to the differences among the models for this forecast horizon; the range of root mean-squared errors is never that large and in the recent period is only about two-tenths of a percentage point.

The two-year ahead forecasts in table 3 present a similar picture. No individual model does well across all the sub-samples, although the diffusion model does perform reasonably well. The naive model does surprisingly well after 1985. Indeed, over the entire 1985–2003 period it is only one-tenth of a percentage point worse than the best individual model for this period, the diffusion model. The range of forecast errors is, as expected, a little larger for the two-year ahead forecasts, but still quite small.

Overall, table 3 indicates that no individual model consistently beats the naive model, and when one model does do better, the gains are small. We conclude that the natural rate, activity, diffusion, combination, and indicator models are not robust inflation forecasting frameworks.

Finally, it is interesting to note the relative performance of the combination, diffusion, and indicator models. Recall that these models involve using many indicators to forecast inflation, but do so in different ways. At the one-year horizon, there is little to choose between the models. Indeed the difference between the models is always less than one-tenth of a percentage point (not shown). At the two-year horizon, the diffusion model consistently outperforms the other two models except for the most recent period. Here the gains are more substantial (also not shown). For example, the diffusion model is superior to the indicator model by over 1 percentage point in the pre-1985 period and superior to the combination model by eight-tenths of a percentage point. In the post-1985 period the gains are about two-tenths and one-tenth of a percentage point, respectively.

The gains to combining forecasts

We now consider what happens when we combine the information in the forecasts from the various models. That is, we add to the list of models compared with the naive model the equally weighted and optimally weighted meta models. For good measure, we throw meta models based on rolling regressions into the mix. These are indicated in the table by the term "rolling." The meta models are indicated by bold type in table 3. Since the optimally weighted models require a sample of forecasts to compute the weights, we only include these models in the mix after 1985. The meta models consist of the naive, natural rate, indicator, activity, diffusion, and combination models.

Notice that for both forecast horizons, the meta models generally outperform the individual models. Moreover, there is always a meta model that outperforms the naive model no matter which sub-sample we consider. Of special note is that it is possible to beat the naive model in the challenging 1985–92 period. Still, overall, the gains over the naive model are modest. Using the rolling regression approach provides some additional gain. At the one-year horizon, the regression strategy for computing weights seems to do better than just averaging the forecasts, but at the two-year horizon the opposite is true.

Is there evidence of a robust model here? Looking at the different sample periods and forecast horizons, it seems that the rolling optimally weighted model consistently outperforms the naive model and is near the top of the performance lists for the one-year horizons. The rolling equally weighted model is a very good performer at the two-year horizon. In both cases, when the model is not at the top of the performance list, it is within one-tenth of a percentage point of the top model and usually much less than that. The gains relative to the naive model are small in the 1985–92

TABLE 3

Monthly RMSE ranking, including meta and rolling models: One-year and two-year ahead forecasts

A. 1-year ahead forecasts	1977–84	1985–92	1993–2000	2001–03	1985–2003
	Natural rate	**Optimally weighted**	**Rolling optimally weighted**	**Rolling optimally weighted**	**Rolling optimally weighted**
	Equally weighted	**Rolling equally weighted**	**Rolling equally weighted**	**Optimally weighted**	**Rolling equally weighted**
	Rolling equally weighted	**Rolling optimally weighted**	**Optimally weighted**	Natural rate	**Optimally weighted**
	Naive	Naive	**Equally weighted**	**Rolling equally weighted**	**Equally weighted**
	Activity	**Equally weighted**	Diffusion	**Equally weighted**	Naive
	Indicators	Combination	Naive	Naive	Combination
	Combination	Autoregression	Activity	Combination	Autoregression
	Autoregression	Indicators	Natural rate	Autoregression	Diffusion
	Diffusion	Diffusion	Combination	Diffusion	Natural rate
	Natural	Rate	Autoregression	Indicators	Indicators
		Activity	Indicators	Activity	Activity
Summary statistics					
Best RMSE	1.03	0.50	0.33	0.38	0.42
Worst RMSE – Best RMSE	0.49	0.39	0.23	0.29	0.28
| Naive RMSE – Best RMSE |	0.20	0.02	0.11	0.12	0.06
Average inflation	6.48	3.84	1.87	1.57	2.65

B. 2-year ahead forecasts	1977–84	1985–92	1993–2000	2001–03	1985–2003
	Rolling equally weighted	**Rolling equally weighted**	**Optimally weighted**	**Equally weighted**	**Rolling equally weighted**
	Equally weighted	**Rolling optimally weighted**	**Rolling optimally weighted**	Natural rate	**Rolling optimally weighted**
	Diffusion	Naive	**Rolling equally weighted**	**Rolling optimally weighted**	**Optimally weighted**
	Naive	Diffusion	**Equally weighted**	**Optimally weighted**	Diffusion
	Natural rate	**Optimally weighted**	Diffusion	Combination	Naive
	Activity	**Equally weighted**	Activity	**Rolling equally weighted**	**Equally weighted**
	Combination	Combination	Naive	Naive	Combination
	Autoregression	Autoregression	Indicators	Autoregression	Autoregression
	Indicators	Indicators	Combination	Indicators	Indicators
		Activity	Natural rate	Diffusion	Activity
		Natural rate	Autoregression	Activity	Natural rate
Summary statistics					
Best RMSE	1.62	0.60	0.39	0.30	0.54
Worst RMSE – Best RMSE	1.32	0.87	0.45	0.47	0.57
| Naive RMSE – Best RMSE |	0.50	0.12	0.35	0.25	0.16
Average inflation	6.48	3.84	1.87	1.57	2.65

Notes: RMSE is root mean-squared error. Meta models are in bold above and include the following individual models: naive, activity, diffusion, combination, natural rate, and indicators.

period, but there are gains. Since 1993, the best meta-models beat the naive model by about one-tenth of a percentage point at the one-year horizon and two-and-a-half-tenths at the two-year horizon. This latter advantage is not insubstantial given that inflation over this period is on average less than 2 percent.

The robust models

Since 1985, the most robust models seem to be the rolling equally weighted and rolling optimally weighted models. It is instructive to study these models a little more.

Cumulative forecast errors

Figures 1 and 2 display cumulative squared forecast errors for the rolling optimally weighted model and the naive model for the one-year and two-year horizons. Figures 3 and 4 (p. 22) are similar, but with the rolling equally weighted and naive models. The vertical lines in these figures indicate the boundaries of the sample periods we consider. To interpret these figures, note that differences in performance are indicated by differences in the slopes of the lines. The model with the flatter line is performing better than the other model over the particular period in which the line is flatter. When one line is below another at a particular date, the model associated with that line has performed better in an RMSE sense up to that date. Note that, due to the need to have data to compute the weights, the figures for the rolling optimally weighted model begin in 1985.

Consider the rolling optimally weighted model first. For the one-year horizon there is little to choose between this model and the naive model in the 1985–92 period. Differences emerge after 1993, but these are concentrated in 1994 and 1995. Additional gains relative to the naive model appear in 2003, though. For the two-year horizon the differences are more substantial, but the overall impression is similar. The location of when the largest gains appear is interesting, since these correspond to periods in which inflation was increasing.

The figures for the rolling equally weighted model present a similar picture for the post-1985 period. The pre-1985 observations are particularly interesting. These illustrate the fact that most of the gains relative to the naive model are in the period before 1985. We can see this in the distance between the two lines in the figures, which does not get much wider after 1985.

Model weights

Figures 5 and 6 (pp. 23–24) display the evolution of the weights underlying the rolling optimally weighted model for the one-year and two-year horizons, respectively. Recall that these weights are based on regressing actual inflation on forecasts from six models, the naive, activity, natural rate, indicator, combination, and diffusion models. The individual models are estimated using rolling regressions, but the weights are based on forecasts for the entire available sample.

Figure 5 shows that for much of the sample all the models get a non-trivial weight for the one-year horizon. Except for the early part of the sample, the weights have not changed that much. Still, their time paths provide some interesting insight into the evolution of the economy. For example, the natural rate model has declined in importance over the sample. Nonetheless, it still gets a large weight. The weight on the naive model has grown over the sample. The activity, diffusion, and combination models get negative weights.[15] Figure 6 indicates that forecasting the two-year horizon involves using the models differently. The natural rate model gets much less weight, and for much of the sample the activity and indicator models get very small weights. Consistent with their individual performances (see table 3), the naive and diffusion models get large weights.

Quarterly data

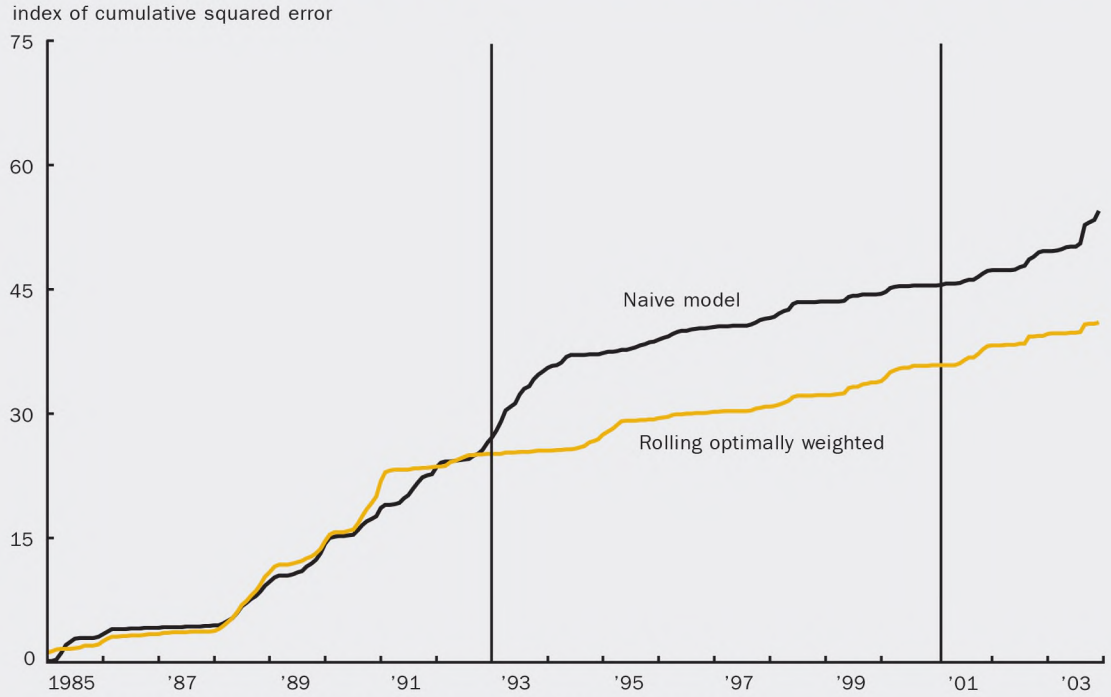
Now, we briefly summarize our findings with quarterly data. To conserve space we do not display our findings. Our purpose here is twofold. We want to know whether averaging the forecasts obtained by different forecasting procedures also improves forecasts at the quarterly frequency. We also want to understand whether adding quarterly data to the analysis that are not available at the monthly frequency improves the quality of the forecasts. The new data include data from the *National Income and Product Accounts*, the output gap, and data on productivity and costs (see the appendix for a list of the specific series).

Regarding the first question, we find that the basic principle of averaging different forecasts also yields forecasting benefits at the quarterly frequency. Indeed the same meta models that show promise at the monthly frequency are also among the most robust at the quarterly frequency when we include the additional quarterly data.[16] With one exception, these models improve on the naive forecast over all sub-samples and both forecast horizons we consider. The exception is in the 1985–92 period for the one-year horizon, in which no model is superior to the naive model.

Incorporating the additional data leads to mixed results. We use the third month in each quarter to compare a given monthly model with its quarterly counterpart. When we do this and compare corresponding monthly and quarterly models, we find little evidence

FIGURE 1

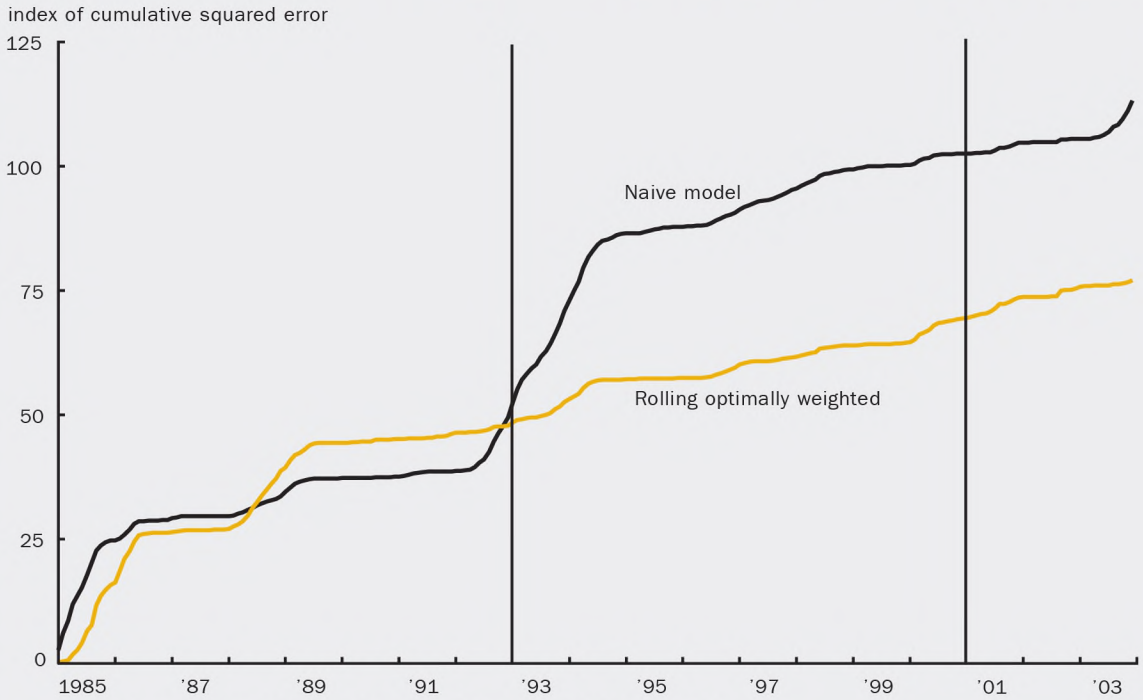
Cumulative squared errors at the 1-year horizon: Naive and rolling optimally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 2

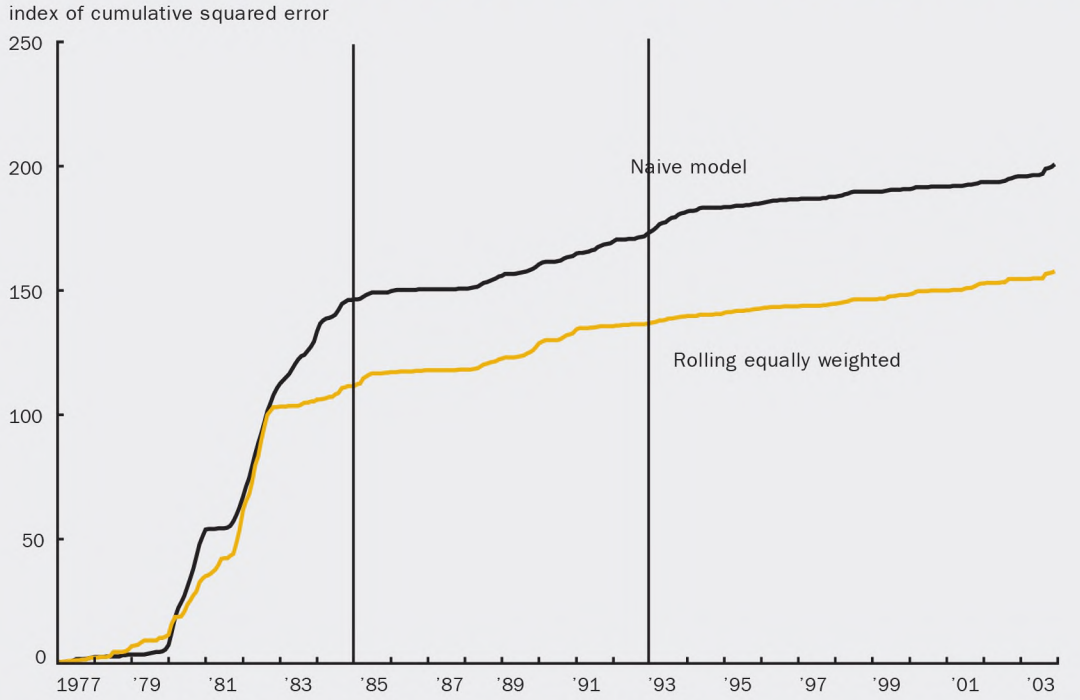
Cumulative squared errors at the 2-year horizon: Naive and rolling optimally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 3

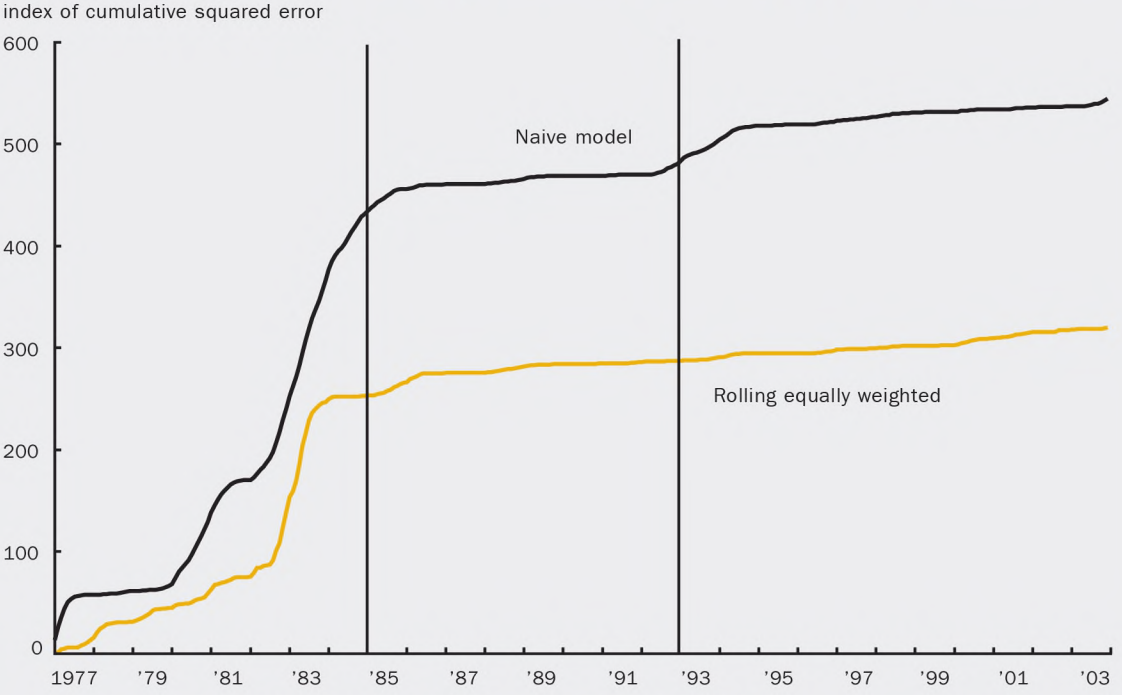
Cumulative squared errors at the 1-year horizon: Naive and rolling equally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 4

Cumulative squared errors at the 2-year horizon: Naive and rolling equally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 5

Regression weights for rolling forecasts, 1-year horizon

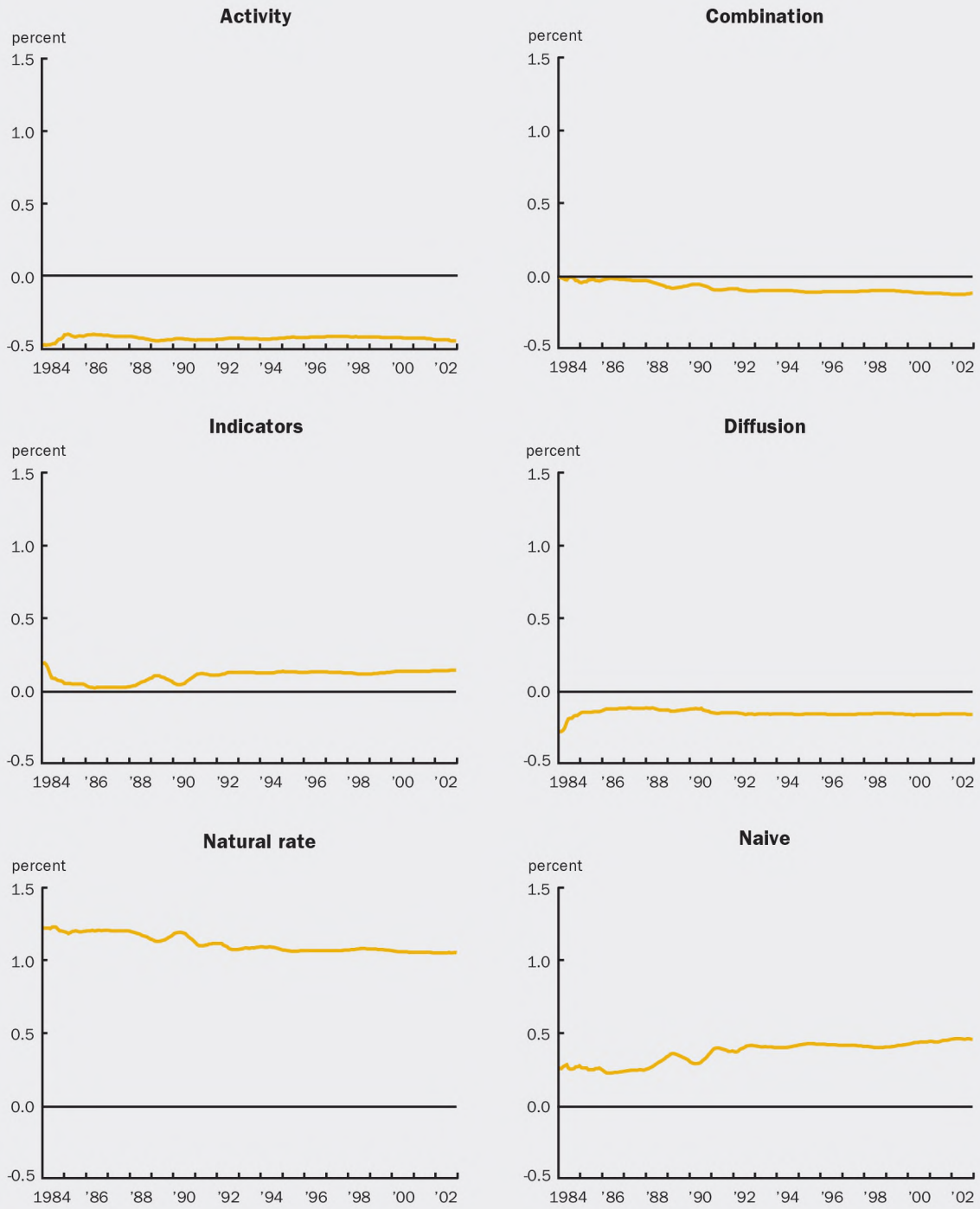
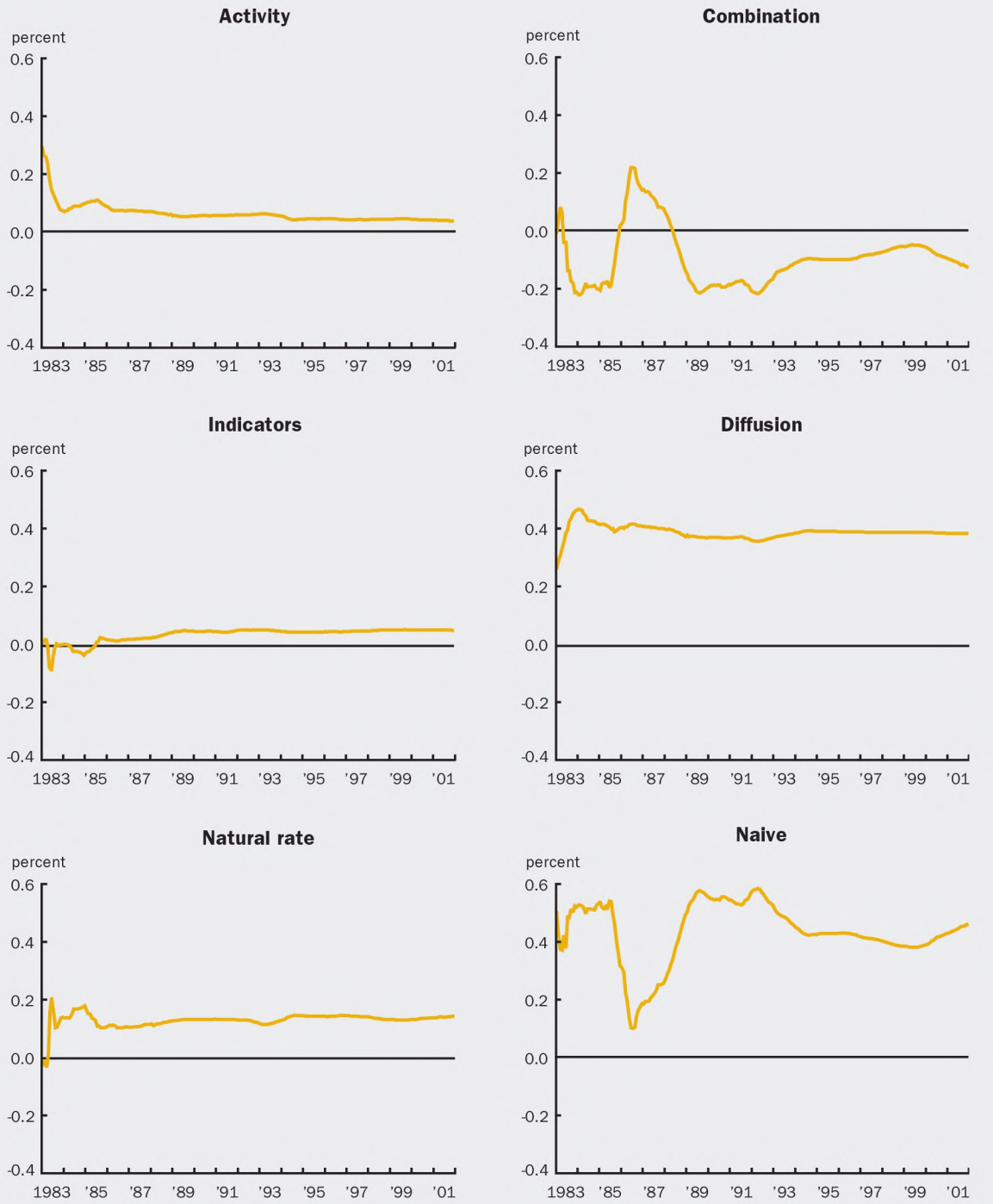


FIGURE 6

Regression weights for rolling forecasts, 2-year horizon



that the additional data improve the forecasts. In particular, there is not a consistent pattern of improvement with the quarterly models and when there is improvement it is typically much less than one-tenth of a percentage point. Sometimes the quarterly models are worse. One model does show consistent improvement at the quarterly frequency—the rolling optimally weighted model. This model does well at the two-year horizon, improving over its monthly counterpart by about one-tenth of a percentage point in all sub-samples after 1985.[17]

In a departure from the monthly analysis, a non-meta model shows up in the list of robust models when we incorporate the additional data. This model is the rolling output gap model, which we could not examine at the monthly frequency because gross domestic product data are only available quarterly. When the output gap model is estimated using the rolling procedure, it is the best performing model over 1977–84 and 1985–2003 and performs better than the naive model in all the sub-samples we consider when forecasting two years ahead. This model does not do as well forecasting at the one-year horizon. In particular, it is outperformed by the rolling optimally weighted model over all the sub-samples. Still, the fact that such a simple model does so well at forecasting two years ahead is interesting and deserves further study.[19]

Taking all the evidence into account, it seems reasonable to conclude that the quarterly data do not add much to forecast performance. Two exceptions are when the additional data are incorporated into the rolling output gap model and the rolling optimally weighted model, both of which perform well at the two-year horizon.

Conclusion

We have found that a robust forecast of the magnitude of inflation can be obtained by combining the forecasts of several models that incorporate the information in the available data in different ways. This suggests that a useful approach to building a reliable statistical forecasting framework is to be eclectic with respect to both the data used to formulate a forecast and the models used to incorporate the data into a forecast. Relying on a small number of inflation indicators and one forecasting model is not a good idea.

Having drawn this conclusion, we must note two caveats.[18] The most obvious caveat is that the conclusion we have just stated sows the seeds of future failure. We have concluded that one must not rely on a particular model, yet we have essentially described a particular model. While we realize the circularity of our conclusion, we would rather interpret our findings as suggesting that combining the forecasts from models that include the data in different ways is the main lesson to be learned. That is, we do not put a lot of weight on the particular models we worked with. We also want to emphasize the limitations of the kinds of forecasting models studied in this article. Clearly, these models are not structural and, therefore, are inadequate for assessing the impact of systematic changes in policy. This is what fully articulated general equilibrium economic models, which account for behavioral responses to policy changes, are for. However, such models, while beginning to be used at central banks, are still inadequate for the everyday needs of policymakers. The forecasting models discussed here have their uses and probably will continue to be popular for some time to come. Principally, these models are useful for understanding what current inflation expectations are. Since the past actions of the Fed are embedded in the coefficients, the models take into account "typical" Fed responses to current conditions. For these reasons, inflation forecasts serve as a useful benchmark for policymakers assessing the current stance of monetary policy. This article has shown that such forecasts can be improved reliably by taking into account information in variables other than inflation.

NOTES

[1]See, for example, Sims (2002) and Stock and Watson (2002). Fisher, Liu, and Zhou (2002) document that the failure of Phillips curve models after 1985 is essentially due to an especially poor performance in the 1985–92 period.

[2]One might view equation 1 as an odd choice to base inflation forecasts on since it involves *changes* of inflation rather than *levels* of inflation. The reason we use this equation is because it performs better than models based on the level of inflation. This reflects the fact that 12-month inflation is an extremely persistent variable, so that its level does not change much over short periods.

[3]Specifically, we use the Bayes information criterion (BIC) to select the number of lags. Intuitively, BIC selects the number of lags to improve the fit of the model without increasing by too much the sampling error in the lag coefficients.

[4]Another way to forecast inflation would be to formulate a vector autoregression in the level or change in one-month inflation and the indicator variables and project this system forward J periods from date T. Such a forecast would yield superior results if the vector autoregression were correctly specified. The conventional wisdom is that the direct approach taken here is in practice better. Marcellino, Stock, and Watson (2004) show that for many variables, but not for inflation, this conventional wisdom is apparently false. We have explored the "multi-step iterated forecasts" described in Marcellino, Stock, and Watson (2004) and concur with their finding that this approach is a poor forecasting strategy for inflation.

[5]To estimate the natural rate, we use a filter applied to the time series of unemployment available at the time of the forecast. The particular filter we use is called a *band-pass filter*. This is designed to isolate particular frequencies of the data. We use it to isolate "long-run" or low frequency fluctuations in the unemployment rate. Specifically, we focus on fluctuations of period (inversely related to the frequency) 12 years or greater. The particular implementation of the band-pass filter we use is the one due to Christiano and Fitzgerald (1999).

[6]The index methodology was proposed by Stock and Watson (1999, 2002). For more details on the CFNAI, see www.chicagofed.org/economic_research_and_data/cfnai.cfm.

[7]Technically, we compute the first six principal components of the 145 variables.

[8]The median of six forecasts is the average of the third and fourth ranked forecasts. We explored other ways of choosing among the six models, including using the mean and using the best out-of-sample forecasting performance (this is described later) up to the date of the forecast. These other ways of summarizing the forecasts performed similarly to the approach taken here.

[9]The word "meta" is often used to describe an analysis that synthesizes research results obtained using different approaches to a question. By this definition, the diffusion, combination, and indicator models might also be considered meta models. We prefer not to use this descriptor to classify these models since they combine the information from forecasts that, except for the indicators used, are based on the same forecasting strategy.

[10]We use this measure of inflation since it plays a prominent role in FOMC discussions.

[11]Compiling the data that were available at a particular point in time is a daunting task. A real-time dataset is available from the Philadelphia Fed. Unfortunately this dataset has a limited number of variables and excludes many that might be useful for forecasting inflation.

[12]Data revisions are a problem for the naive and autoregression models since the price index we use, the PCE deflator, is subject to revisions.

[13]Comparisons of models based on RMSE are subject to sampling variability and consequently subject to error. In principle, we could use Monte Carlo methods to assess the magnitude of this error. However, this would require specifying an underlying data-generating process for all the variables in our analysis (more than 150 of them). This sampling error should be kept in mind when interpreting the results. See Clark and McCracken (2001) and the references they cite for a useful discussion of some of the issues involved in assessing the statistical difference in the accuracy of forecasts.

[14]For another discussion of this point, see Cecchetti, Chu, and Steindel (2000).

[15]In principle there is nothing wrong with a negative weight. Conditional on all the other forecasts, a forecast of an increase in inflation from a model with a negative weight is a signal that the other models combined are forecasting an increase in inflation that is too big or a decrease in inflation that is not big enough, relative to past experience. If the model did not provide information about inflation, then it would get a zero weight.

[16]When computing the weighted forecasts at the quarterly horizon, we add the forecasts of the output gap model to the list of forecasts that are averaged.

[17]We also examine the impact of just averaging the monthly data to convert it to the quarterly frequency. When we do this, we find little evidence that monthly noise is a significant source of forecast error since there is not a consistent pattern of improvement in the quarterly models and when there is improvement it is typically much less than one-tenth of a percentage point.

[18]Another important caveat involves the use of rolling regressions. Sargent (1999) argues that the rise of inflation during the 1960s and 1970s and the subsequent decline can be explained by a process of the Fed learning and forgetting about its ability to exploit a perceived trade-off between inflation and unemployment. This analysis suggests a potential problem with using the rolling regression framework, because it may lead to a recurrence of the rise of inflation in the 1960s and 1970s. However, as Sargent (1999, p. 134) points out, a credible commitment by the Fed to low inflation should prevent such a recurrence. Under this view, there is no problem with using the rolling regression approach to forecasting.

[19]See Clark and McCracken (2004) for a recent analysis of the predictive content of the output gap for inflation.

DATA APPENDIX

Monthly data: 1967:01–2003:12[a]

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Activity	log 1st diff	le		Civilian employment: Sixteen years & over: 16 yr + (SA, 000s)	usecon	
Activity	1st diff	lrm25		Civilian unemployment rate: Men, 25–54 years (SA, %)	usecon	
Activity	1st diff		LCUN = a0m005	Average weekly initial claims unemployment insurance (SA, 000s)	bci	
Activity, diffusion, combination	log 1st diff	cbhm		Personal consumption expenditures (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	cdbhm		Personal consumption expenditures: Durable goods (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	cnbhm		Personal consumption expenditures: Nondurable goods (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	csbhm		Personal consumption expenditures: Services (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	ypdhm		Real disposable personal income (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff		CONSTPV = cpv –cpvr	Value of public construction put in place (SAAR, chained $mil.)	usecon	
Activity, diffusion, combination	log 1st diff		CONSTPU = cpg	Value of private construction put in place (SAAR, chained $mil.)	usecon	
Activity, diffusion, combination	log	hsm		Manufacturers' shipments of mobile homes (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hst		Housing starts (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstmw		Housing starts: Midwest (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstne		Housing starts: Northeast (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hsts		Housing starts: South (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstw		Housing starts: West (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log 1st diff	ip		Industrial Production Index (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip51		Industrial Production: Consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip511		Industrial Production: Durable consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip512		Industrial Production: Nondurable consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip521		Industrial Production: Business equipment (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip53		Industrial Production: Materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip531		Industrial Production: Durable goods materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip532		Industrial Production: Nondurable goods materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip54		Industrial Production: Nonindustrial supplies (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipb0		Industrial Production: Mining (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipfp		Industrial Production: Final products (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmdg		Industrial Production: Durable goods [NAICS] (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmfg		Industrial Production: Manufacturing [SIC] (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmnd		Industrial Production: Nondurable manufacturing (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	iptp		Industrial Production: Final products and nonindustrial supplies (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	iputl		Industrial Production: Electric and gas utilities (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	laconsa		All employees: Construction (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	ladurga		All employees: Durable goods manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lafirea		All employees: Financial activities (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lagooda		All employees: Goods-producing industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lagovta		All employees: Government (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lamanua		All employees: Manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	laminga		All employees: Mining (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lanagra		All employees: Total nonfarm (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	landura		All employees: Nondurable goods manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lapriva		All employees: Total private industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lartrda		All employees: Retail trade (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	laserpa		All employees: Service-providing industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	LASRVSA = lainfoa + lapbsva + laeduha + laleiha + lasrvoa		All employees: Aggregate of categories	usecon	
Activity, diffusion, combination	log 1st diff	LATPUTA = lattula – lawtrda – lartrda		All employees: Aggregate of categories	usecon	
Activity, diffusion, combination	log 1st diff	lena		Civilian employment: Nonagricultural Industries: 16yr + (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lhelpr		Ratio: Help-wanted advertising in newspapers/Number unemployed (SA)	usecon	
Activity, diffusion, combination	1st diff	lomanua		Average weekly hours: Overtime: Manufacturing (SA, Hrs)	usecon	
Activity, diffusion, combination	1st diff	lrmanua		Average weekly hours: Manufacturing (SA, Hrs)	usecon	
Activity, diffusion, combination	level	napmc		ISM Mfg: PMI Composite Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmei		ISM Mfg: Employment Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmii		ISM Mfg: Inventories Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmni		ISM Mfg: New Orders Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmoi		ISM Mfg: Production Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	log 1st diff	rsdh		Real retail sales: Durable goods (SA, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		RSH = rsh + rsh2	Retail sales: Retail trade (SA, Spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff	rsnh		Real retail sales: Nondurable goods (SA, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMDH = timdh + timdh2	Real inventories: Mfg: Durable goods industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMH = timh + timh2	Real manufacturing & trade inventories: Mfg industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMNH = timnh + timnh2	Real mfg inventories: Nondurable goods industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIRH = tirh + tirh2	Real inventories: Retail trade industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TITH = tith + tith2	Real manufacturing & trade inventories: Industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIWH = tiwh + tiwh2	Real inventories: Merchant wholesale trade industries (SA, EOP, spliced, chained 2000$mil.)	usecon	

DATA APPENDIX *(continued)*

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Activity, diffusion, combination	1st diff		TRMH = trmh + trmh2	Real inventories/sales ratio: Manufacturing industries (SA, spliced, chained 2000$)	usecon	
Activity, diffusion, combination	1st diff		TRRH= trrh + trrh2	Inventories/sales ratio: Retail trade industries (SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	1st diff		TRTH= trth + trth2	Real manufacturing & trade: Inventories/sales ratio (SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	1st diff		TRWMH=trwmh + trwmh2	Inventories/sales ratio: Merchant wholesale trade industries(SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	log 1st diff		TSMDH= tsmdh + tsmdh2	Real sales: Mfg: Durable goods industries(SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSMH= tsmh + tsmh2	Real sales: Manufacturing industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSMNH= tsmnh + tsmnh2	Real sales: Mfg: Nondurable goods industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSTH= tsth + tsth2	Real manufacturing & trade sales: All industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMDH= tswmdh	Real sales: Merchant wholesalers: Durable goods inds. (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMH= tswmh + twsmh2	Real sales: Merchant wholesale trade industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMNH= tswmnh + tswmnh2	Real sales: merchant wholesale: Nondurable goods inds. (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff	ypltpmh		Real personal income less transfer payments (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff		CDVHM = cdvhm + cdvh	PCE: Durable goods: Motor vehicles and parts (SAAR, spliced and interpolated, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		MDOQ = a0m007	Manufacturers' new orders: Durable goods (SA, chained 2000$mil.)	bci	
Activity, diffusion, combination	log 1st diff		MOCGMC = a0m008	Manufacturers' new orders: Consumer goods & materials (SA, 1982$mil.)	bci	
Activity, diffusion, combination	log 1st diff		MOCNC = a0m027	Manufacturers' new orders: Nondefense capital goods (SA, 1982$mil.)	bci	
Activity, diffusion, combination, indicators (3)	log	hpt		New private housing units authorized by building permit (SAAR, units in 000s)	usecon	
Activity, diffusion, combination, indicators (2)	1st diff	cumfg		Capacity utilization: Manufacturing [SIC] (SA, % of capacity)	usecon	
Activity, diffusion, combination, indicators (2)	log 1st diff	lhelp		Index of help-wanted advertising in newspapers (SA, 1987=100)	usecon	
Activity, diffusion, combination, indicators	1st diff	lr		Civilian unemployment rate: 16yr + (SA, %)	usecon	
Activity, indicators (2)	level	napmvdi		ISM: Mfg: Vendor Deliveries Index (SA, 50+ = Econ Expand)	usecon	
Diffusion, combination	level	cexp		University of Michigan: Consumer expectations (NSA, 66Q1=100)	usecon	
Diffusion, combination	level	lu0		Civilians unemployed for less than 5 weeks (SA, 000s)	usecon	
Diffusion, combination	level	lu15		Civilians unemployed for 15–26 weeks (SA, 000s)	usecon	
Diffusion, combination	level	lu5		Civilians unemployed for 5–14 weeks (SA, 000s)	usecon	
Diffusion, combination	level	luad		Average (Mean) duration of unemployment (SA, weeks)	usecon	
Diffusion, combination	level	lut15		Civilians unemployed for 15 weeks and over (SA, 000s)	usecon	
Diffusion, combination	level	lut27		Civilians unemployed for 27 weeks and over (SA, 000s)	usecon	
Diffusion, combination	log 2nd diff	faram		Adjusted monetary base (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	faran		Adjusted nonborrowed reserves of depository institutions (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	faranp		Adjusted nonborrowed reserves plus extended credit (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	farat		Adjusted reserves of depository institutions (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	farmsr		Adj. monetary base including deposits to satisfy clearing balance contracts (SA, $bil.)	usecon	
Diffusion, combination	log 2nd diff	fm1		Money stock: M1 (SA, $bil.)	usecon	
Diffusion, combination	log 1st diff	fm2c		Real money stock: M2 (SA, chained 2000$bil.)	usecon	
Diffusion, combination	log 2nd diff	fm3		Money stock: M3 (SA, $bil.)	usecon	
Diffusion, combination[a]	log 1st diff	fxtwb[a]		Nominal broad trade-weighted exchange value of US$ (JAN 97=100)	usecon	
Diffusion, combination	log 1st diff	fxuk		Foreign exchange rate: United Kingdom (US$/Pound)	usecon	
Diffusion, combination	1st diff	faaa		Moody's seasoned Aaa corporate bond yield (% p.a.)	usecon	
Diffusion, combination	1st diff	fbaa		Moody's seasoned Baa corporate bond yield (% p.a.)	usecon	
Diffusion, combination	level		DAAA = faaa – ffed	Moody's seasoned Aaa corporate bond yield – fed funds rate(% p.a.)	usecon	
Diffusion, combination	level		DBAA = fbaa –ffed	Moody's seasoned Baa corporate bond yield – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level	sdy5comm		S&P: Composite 500, dividend yield (%)	usecon	
Diffusion, combination	log 1st diff	sp500		Stock Price Index: Standard & Poor's 500 Composite (1941–43=10)	usecon	
Diffusion, combination	level	spe5comm		S&P: 500 Composite, P/E ratio, 4-qtr trailing earnings	usecon	
Diffusion, combination	log 1st diff	spny		Stock Price Index: NYSE Composite (Avg, Dec. 31, 2002=5000)	usecon	
Diffusion, combination	log 1st diff	spspi		Stock Price Index: Standard & Poor's 400 Industrials (1941–43=10)	usecon	
Diffusion, combination	1st diff	ftbs3		3-month Treasury bills, secondary market (% p.a.)	usecon	
Diffusion, combination	1st diff	ftbs6		6-month Treasury bills, secondary market (% p.a.)	usecon	
Diffusion, combination	level		DTBS03 = ftbs3 – ffed	3-month Treasury bills – fed funds rate, (% p.a.)	usecon	
Diffusion, combination	level		DTBS06 = ftbs6 – ffed	6-month Treasury bills – fed funds rate (% p.a.)	usecon	
Diffusion, combination	1st diff	fcm1		1-year Treasury bill yield at constant maturity (% p.a.)	usecon	
Diffusion, combination	1st diff	fcm5		5-year Treasury note yield at constant maturity (% p.a.)	usecon	
Diffusion, combination	level		DCM1 = fmc1 – ffed	1-year Treasury bill yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level		DCM5 = fmc5 –ffed	5-year Treasury note yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level		DCM10 = fcm10 –ffed	10-year Treasury note yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	log 2nd diff	sp1000		PPI: Crude materials for further processing (SA, 1982=100)	usecon	
Diffusion, combination	log 2nd diff	sp3100		PPI: Finished consumer goods (SA, 1982=100)	usecon	
Diffusion, combination	log 2nd diff	pcua		CPI-U: Apparel (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcucc		CPI-U: Commodities (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcuccd		CPI-U: Durables (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcucs		CPI-U: Services (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcum		CPI-U: Medical care (SA, 1982–84=100)	usecon	

DATA APPENDIX *(continued)*

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Diffusion, combination	log 2nd diff	pcuslf		CPI-U: All items less food (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcuslm		CPI-U: All items less medical care (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcusls		CPI-U: All items less shelter (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcut		CPI-U: Transportation (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	jcdm		PCE: Durable goods: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcm		PCE: Personal consumption expenditures: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcnm		PCE: Nondurable goods: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcsm		PCE: Services: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	leconsa		Avg hourly earnings: Construction (SA, $/Hr)	usecon	
Diffusion, combination	log 2nd diff	lemanua		Avg hourly earnings: Manufacturing (SA, $/Hr)	usecon	
Diffusion, combination	1st diff		FCLQ = a0m101	Commercial & industrial loans outstanding (EOP, SA, chained 2000$mil.)	bci	
Diffusion, combination, indicators (5)	log 2nd diff	fm2		Money stock: M2 (SA, $bil.)	usecon	
Diffusion, combination, indicators (5)	1st diff	fcm10		10-year Treasury note yield at constant maturity (% p.a.)	usecon	
Diffusion, combination, indicators	1st diff	ffed		Federal funds [effective] rate (% p.a.)	usecon	
Diffusion, combination, indicators (4)	log 2nd diff	sp2000		PPI: Intermediate materials, supplies, and components (SA, 1982=100)	usecon	
Diffusion, combination, indicators (4)	log 2nd diff	sp3000		PPI: Finished goods (SA, 1982=100)	usecon	
Diffusion, combination, indicators (4)	level	napmpi		ISM: Mfg: Prices Index (NSA, 50+ = Econ Expand)	usecon	
Indicators (1) log	1st diff	zlead		Composite Index of 10 Leading Indicators (1996=100)	bci	
Indicators (3) log	1st diff		CPC = CONSTPV + CONSTPU	New construction put in place (SAAR, 2000$mil.)	usecon	
Indicators (3) log	1st diff	hn1us		New single-family houses sold: United States (SAAR, 000s)	usecon	
Indicators (1) log	1st diff	chm		Personal consumption expenditures (SAAR, chained 2000$mil.) (spliced from usna96 before 1990)	usna	
Indicators (1)	level	swxli2		Stock and Watson nonfinancial leading index %	usecon	
Indicators (5) level			CM03CM01 = fcm3 – fmc1	3-year/1-year T-bill spread	usecon	
Indicators[b] (5)	level	fxtwm[b]		Nominal trade-weighted exch value of US$/major currencies (MAR 73=100)	usecon	
Indicators (5)	log 1st diff		PZGLD = pzgld + mgold + fgold	Cash prices: gold, Handy & Harman Base Price (avg, spliced, $/Troy oz)	weekly	COMEX, FSC
Indicators (5)	log 1st diff		PZSIL	Cash price: silver, troy oz, Handy & Harman Base Price (avg, $/troy oz)	weekly	FSC
Indicators (4) log	1st diff	pzall		KR-CRB Spot Commodity Price Index: All commodities	usecon	
Indicators[c] (3) log	1st diff	spwpc[c]		SPOT COMMODITY PRICE - PLYWOOD, CROWS (PUIWMWPC_N.WT)		FAME
Indicators (4) log	1st diff		PFALL	KR-CRB Futures: All commodities (avg, 1967=100) weekly		BCRB
Indicators[d] (4) log	1st diff	pzdalu[d]		Aluminum ingot producer price: Delivered Midwest (avg, cents/lb)	weekly	
Indicators (4) log	1st diff	p101		PPI: Iron and steel (NSA, 1982=100)	usecon	
Indicators (4) log	1st diff	ueg		CPI-U: Energy (SA, 1982–84=100)	cpidata	
Natural rate	band-pass filtered	UGAP		Unemployment gap constructed from Perry-weighted unemployment rate	empl	
Prices	log 2nd diff	jcxfem		PCE less food and energy: Price Index (SA) (2000=100)	usna	

COMEX	http://www.wrenresearch.com.au/downloads/index.htm	Indicator model groups:
FSC	http://www.webspace4me.net/~blhill2/data/commodities	1: Economic activity
BCRB	http://economic-charts.com/em-cgi/data.exe/crb/crb01	2: Slackness measures
FAME	Federal Reserve Bank of San Francisco website	3: Housing and building activity
		4: Industrial prices
		5: Financial markets

[a]fxtwb begins in 1973:01
[b]fxtwm begins in 1973:01
[c]cspwpc begins in 1979:01
[d]pzdalu begins in 1988:07

Notes: SAAR is seasonally adjusted annual rate, SA is seasonally adjusted, NSA is not seasonally adjusted, NAICS is North American industry classification system, SIC is standard industrial classification, and EOP is end of period.

DATA APPENDIX *(continued)*

Quarterly data: 1967:1–2003:4

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database
Activity	log 1st diff	ch		Real Personal Consumption Expenditures (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	cdh		Real Personal Consumption Expenditures: Durable Goods (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	cnh		Real Personal Consumption Expenditures: Non-Durable Goods (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	csh		Real Personal Consumption Expenditures: Services (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	ih		Real Gross Private Domestic Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	fh		Real Private Fixed	usna
Activity	log 1st diff	fnsh		Real Private Nonresidential Structures	usna
Activity	log 1st diff	fneh		Real Private Nonresidential Equipment & Software	usna
Activity	1st diff	vh		Real Change in Private Inventories (SAAR, Bil. Chn. 2000 $)	usna
Activity	1st diff	xneth		Real Net Exports of Goods & Services (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gh		Real Govt. Consumption Expenditures & Gross Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gfnh		Real Govt. Non-defense Consumption Expenditures & Gross Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	ypdh		Real Disposable Personal Income (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gdpbq		Index of Business Gross Value added	usna
Activity	log 1st diff	fsq		Index of Real Final Sales	usna
Activity, Indicators (1)	log 1st diff	gdph		Real Gross Domestic Product (SAAR, Bil. Chn. 2000 $)	usna
Activity, Indicators (3)	log 1st diff	fnh		Real Private Nonresidential	usna
Activity, Indicators (3)	log 1st diff	frh		Real Private Residential	usna
Diffusion, Combination	log 1st diff	lxba		Business Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbc		Business Sector: Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbr		Business Sector: Real Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbu		Business Sector: Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbn		Business Sector: Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnfn		Non-farm Business Sector: Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxma		Manufacturing Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxmda		Manufacturing Sector Durables: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxmna		Manufacturing Sector Non-durables: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnca		Non-financial Corporations: Output per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncc		Non-financial Corporations: Compensation per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncr		Non-financial Corporations: Real Compensation per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncu		Non-financial Corporations: Unit Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncn		Non-financial Corporations: Unit Non-Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnct		Non-financial Corporations: Total Unit Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		BRULC= lxbu/lxbi	Business Sector: Real Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		FRULC= lxncu/lxnci	Non-financial Corporations: Real Unit Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		BNLRULC= lxbn/lxbi	Business Sector: Real Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		NFNLRULC = lxnfn/lxnfi	Non-farm Business Sector: Real Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		FNLRULC= lxncn/lxnci	Non-financial Corporations: Real Unit Non-Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		FTOTRUC= lxnct/lxnci	Non-financial Corporations: Real Total Unit Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	grt		Government Total Receipts (SAAR, Bil. $)	usna
Diffusion, Combination	log 1st diff	get		Government Total Expenditures (SAAR, Bil. $)	usna
Diffusion, Combination	1st diff	gnl		Government Net Lending or Net Borrowing (SAAR, Bil. $)	usna
Diffusion, Combination	log 1st diff	dgdp		GDP Deflator	usna
Diffusion, Combination	log 1st diff	di		Gross Private Domestic Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	df		Private Fixed Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfn		Private Non-residential Fixed Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfns		Private Non-residential Structures: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfne		Private Non-residential Equipment/Software: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfr		Private Residential Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dg		Government Consumption/Gross Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dgfn		Federal Non-Defense Consumption/Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dm		Imports of Goods & Services: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dx		Exports of Goods & Services: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfa		Non-farm Business Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfc		Non-farm Business Sector: Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfr		Non-farm Business Sector: Real Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfu		Non-farm Business Sector: Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff		NFRULC= lxnfu/lxnfi	Non-farm Business Sector: Real Unit Labor Costs (SA,1992=100)	usecon
Output Gap	band-pass filtered		OGAP	Output gap constructed from band-pass filtered Real GDP	usna
Real Unit Labor Cost Gap	band-pass filtered		RGAP	Band-pass filtered version of Non-farm Business Sector Real Unit Labor Costs	usecon
Prices	log 2nd diff	jcxfe		PCE less food and Energy: Price Index (SA) (2000=100)	usna

Indicator Model Groups:
1: Economic Activity
2: Slackness Measures
3: Housing and Building Activity
4: Industrial Prices
5: Financial Markets
6: Productivity and Marginal Cost

REFERENCES

Atkeson, Andrew, and Lee E. Ohanian, 2001, "Are Phillips curves useful for forecasting inflation?," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol. 25, No. 1, Winter, pp. 2–11.

Cecchetti, Stephen G., Rita S. Chu, and Charles Steindel, 2000, "The unreliability of inflation indicators," *Current Issues in Economics and Finance*, April, Vol. 6, No. 4.

Christiano, Lawrence J., and Terry Fitzgerald, 1999, "The band-pass filter," National Bureau of Economic Research, working paper, No. 7257, July.

Clark, Todd E., and Michael W. McCracken, 2004, "The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence," manuscript.

_____, 2001, "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, Vol. 105, November, pp. 85–110.

Fisher, Jonas D. M., Chin Liu, and Ruilin Zhou, 2002, "When can we forecast inflation?," *Economic Perspectives*, Federal Reserve Bank of Chicago, First Quarter, pp. 30–42.

Marcellino, Massimiliano, James H. Stock, and Mark Watson, 2004, "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," manuscript.

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

Sims, Christopher, 2002, "The role of models and probabilities in the monetary policy process," *Brookings Papers on Economic Activity*, Vol. 2, pp.1–62.

Stock, James H., and Mark Watson, 2003, "Forecasting output and inflation: The role of asset prices," *Journal of Economic Literature*.

_____, 2002, "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, Vol. 20, No. 2, April, pp. 147–162.

_____, 1999, "Forecasting inflation," *Journal of Monetary Economics*, Vol. 44, No. 2.

CALL FOR PAPERS AND SAVE THE DATE

May 4–6, 2005

41st ANNUAL CONFERENCE ON BANK STRUCTURE AND COMPETITION
FEDERAL RESERVE BANK OF CHICAGO

The Art of the Loan in the 21st Century: Producing, Pricing, and Regulating Credit

The Federal Reserve Bank of Chicago invites the submission of research- and policy-oriented papers for the 41st annual Conference on Bank Structure and Competition to be held May 4–6, 2005, at the Westin Hotel in Chicago. Since its inception, the conference has fostered an ongoing dialogue on current public policy issues affecting the financial services industry. We welcome submissions of high-quality research papers on *all topics* related to financial services and their regulation. As in past years, the program will highlight the conference theme, but will also feature numerous sessions on topics unrelated to the conference theme.

The theme of the 2005 conference focuses on the evolution of lending. For most of the 20th century, lending channels resembled silos: Businesses borrowed at commercial banks, home buyers borrowed at thrift institutions, and consumers borrowed at finance companies and credit unions. Loans were held on-balance-sheet, and the interest income they generated was the fundamental source of lender profits. Lenders had comparative advantages at gathering financial information, in part because they also provided savings, investment, and payments vehicles for their customers. Only the very largest corporations could borrow in public debt markets.

Astonishing advances in information, communications, and financial technologies have drastically altered this lending landscape. Lenders face more competition and borrowers have more choices. Banks play a key role in this mix, but their role is changing. Increasingly, both business borrowers and household borrowers are benefiting from market financing—either directly as clients of investment banks or indirectly via asset securitization. As interest income from these arrangements flows to investors in corporate bonds and asset-backed securities, lenders have become more reliant on fee income from originating and servicing loans, providing back-up credit facilities, arranging loan syndications, and underwriting debt securities.

These ongoing changes have important implications for the structure and performance of the financial industry, for the amount of credit created and its distribution, and for macroeconomic growth and stability. For example, will improvements in information flows and securitization practices erode the need for relationship-based portfolio lending to small businesses, and with it the primary role for small banks? Will the scale economies inherent in automated, transactions-based credit provision make the economics of lending untenable for small institutions? At the other extreme, will increasing financial market efficiency further erode the market for syndicated portfolio lending to large businesses? Are large banking companies consolidating in pursuit of size-based efficiencies or market power? Have commercial and

investment banks already begun to exploit market power by 'tying' underwriting and lending for large corporate customers? Is there an endgame in which a handful of global banking companies dominate international, or even domestic, credit markets?

While new technologies and financial innovations have generally benefited borrowers by reducing the cost of credit and expanding access to credit, there are concerns about unknown social and economic costs. Easier access to both secured and unsecured consumer credit has unleashed the power of financial leverage and enhanced macroeconomic growth, but some critics worry about increased economic volatility due to personal bankruptcies. The transformation of mortgage banking has been accompanied by an increase in homeownership, but worries about discrimination and predation on unsophisticated borrowers continue. The nascent expansion of credit unions into small business lending improves credit access, but the tax advantage enjoyed by these cooperatives enrages community bankers, and some analysts hear an echo of the thrift crisis.

Finally, the evolution of loan underwriting, funding, and distribution has altered the nature of credit risk. As lending sectors consolidate, credit risk may become unduly concentrated at a small number of large institutions. For example, mortgage-related interest rate risk has become concentrated at large government-sponsored enterprises (GSEs), with potential systemic consequences. Similarly, consumer-related credit risk has become concentrated at large, mono-line credit card banks. How effectively can lenders use derivatives and other risk-mitigation tools to offset these risks? Will the new Basel II capital regulations create the proper balance of incentives for risk-taking by institutions that originate, sell, hold, or participate in loans? Do these developments dampen, or exacerbate, the effects of credit markets on business cycles?

In addition to the above issues directly related to the 2005 conference theme, the program committee is interested in evaluating submissions on the following topics:

- Competitive strategies of financial institutions;
- The regulation and riskiness of government-sponsored enterprises (GSEs);
- The Basel II Capital Accord;
- Financial industry consolidation;
- Payments innovations;
- Credit access: fair lending, the Community Reinvestment Act, and predatory lending issues;
- Deposit insurance and safety net reform;
- Measuring, monitoring, and managing bank risk;
- The viability and future role of community banks; and
- Restructuring of financial regulatory agencies.

If you would like to present a paper at the conference, please submit four copies of the completed paper or a detailed abstract (the more complete the paper, the better) with your name, address, affiliation, telephone number, and e-mail address, and those of any co-authors, by December 22, 2004. Correspondence should be addressed to:

Conference on Bank Structure and Competition
Research Department
Federal Reserve Bank of Chicago
230 South LaSalle Street
Chicago, Illinois 60604-1413

Alternatively, you may submit your manuscript and personal information via email to Regina Langston (rlangston@frbchi.org). For additional information contact:

Douglas Evanoff at 312-322-5814,
Robert DeYoung at 312-322-5396, or
Regina Langston at 312-322-5641.

How do banks make money? The fallacies of fee income

Robert DeYoung and Tara Rice

Introduction and summary

"How do banks make money?" is a deceivingly simple question. Banks make money by charging interest on loans, of course. In fact, there used to be a standard, tongue-in-cheek answer to this question: According to the "3-6-3 rule," bankers paid a 3 percent rate of interest on deposits, charged a 6 percent rate of interest on loans, and then headed to the golf course at 3 o'clock.

Like most good jokes, the 3-6-3 rule mixes a grain of truth with a highly simplified view of reality. To be sure, the interest margin banks earn by intermediating between depositors and borrowers continues to be the primary source of profits for most banking companies. But banks also earn substantial amounts of noninterest income by charging their customers fees in exchange for a variety of financial services. Many of these financial services are traditional banking services: transaction services like checking and cash management; safe-keeping services like insured deposit accounts and safety deposit boxes; investment services like trust accounts and long-run certificates of deposit (CDs); and insurance services like annuity contracts. In other traditional areas of banking—such as consumer lending and retail payments—the widespread application of new financial processes and pricing methods is generating increased amounts of fee income for many banks. And in recent years, banking companies have taken advantage of deregulation to generate substantial amounts of noninterest income from nontraditional activities like investment banking, securities brokerage, insurance agency and underwriting, and mutual fund sales.

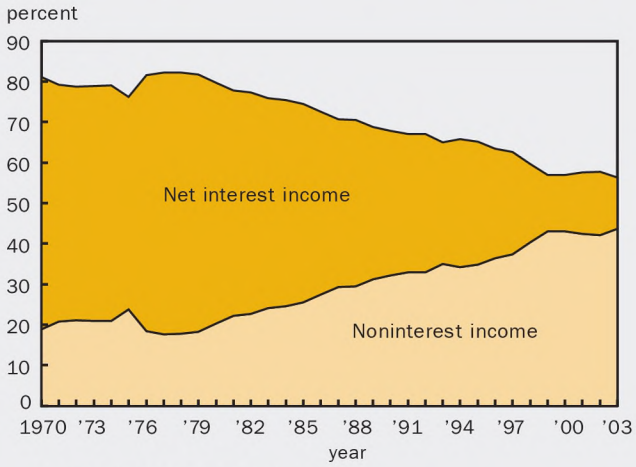
Remarkably, noninterest income now accounts for nearly half of all operating income generated by U.S. commercial banks. As illustrated in figure 1, fee income has more than doubled as a share of commercial bank operating income since the early 1980s. This shift has been larger than most industry experts expected, and we have only recently begun to understand the implications of this shift for the financial performance of banking companies. Only a handful of systematic academic studies have been completed thus far, and those studies have tended to contradict the conventional industry beliefs about noninterest income. Many in the banking industry continue to discount, underestimate, or simply misunderstand the manner in which increased noninterest income has affected the financial performance of banking companies.

This article documents the dramatic increase in noninterest income at U.S. banking companies during the past two decades, the myriad forces that have driven this increase, and the somewhat surprising implications of these changes for the financial performance of commercial banks. We pay special attention to two fundamental misunderstandings about noninterest income at commercial banks. The first is the belief that noninterest income and fee income are more stable than interest-based income. We review the most recent evidence from academic studies that strongly suggest—contrary to the original expectations of many—that increased reliance on fee-based activities tends to increase rather than decrease the volatility of banks' earnings streams. The second misunderstanding is the belief that banks earn noninterest income chiefly from nontraditional, nonbanking activities. We perform some calculations of our own and demonstrate that payment services—one of the most traditional of all

Robert DeYoung is a senior economist and economic advisor and Tara Rice is an economist in the Economic Research Department of the Federal Reserve Bank of Chicago. *The authors thank Carrie Jankowski and Ian Dew-Becker for excellent research assistance and Bob Chakravorti, Cindy Bordelon, and Craig Furfine for helpful comments.*

FIGURE 1

Noninterest and net income as a % of total operating income in U.S. commercial banking, 1970–2003



Note: The two series sum to 100 percent.

banking services—remain the single largest source of noninterest income at most U.S. banking companies.

This is the first of two articles in this issue of *Economic Perspectives* that examine "how banks make money." The companion piece that follows describes the wide diversity of business strategies being used by commercial banking companies—some of which rely disproportionately on activities that generate noninterest income—and compares and contrasts the risk-return profiles of banking companies that employ those strategies.

Noninterest income, deregulation, and technological change

Banks earn noninterest income by producing both traditional banking services and nontraditional financial services. In fact, even before deregulation provided banks with increased opportunities to sell nontraditional fee-based services (say, in the mid-1980s), noninterest income already represented about $1 out of $4 of operating income generated by commercial banks. And the dramatic increase in noninterest income at U.S. banking companies over the past two decades reflects not only a diversification of banks into nontraditional activities, but also a shift in the way banks earn money from their traditional banking activities.

Table 1 organizes selected fee-generating activities into two groups: traditional activities that have always been provided by commercial banks and nontraditional financial services that banks have only recently begun to provide. (This is a selected list of activities for illustrative purposes only and is not meant to cover all fee-based activities.)

The first column in table 1 would have been empty for the years prior to the deregulation of the financial industry. Deregulation opened the door for commercial banks to earn fee income from investment banking, merchant banking, insurance agency, securities brokerage, and other nontraditional financial services. The key deregulation was the Gramm–Leach–Bliley (GLB) Act of 1999, which created a financial holding company (FHC) framework that allowed common ownership of, and formal affiliation between, banking and nonbanking activities. Although GLB was the "big bang" that eliminated most of the Glass–Steagall Act (1933) prohibitions on mixing commercial banking and other financial services, partial deregulation had occurred during the 1980s and 1990s. In the late 1980s the Federal Reserve allowed commercial banks to set up investment banking subsidiaries with limited underwriting powers, and in the mid-1990s the Office of the Comptroller of the Currency granted national banks the power to sell insurance from offices in small towns.

The fees generated by these new, nontraditional activities are uneven across banking companies. On the one hand, investment banking has been a natural addition to the product lines of large banking companies that have large corporate clients. On the other hand, insurance agency has been a good fit for banking companies of all sizes that wish to cross-sell new financial services to their retail (household) clients.

In contrast, the fee-generating activities listed in the second column of table 1 are very traditional banking activities. Banks have always earned noninterest income from their depositors, charging fees on a variety of transaction services (for example, checking and money orders), safe-keeping services (for example, insured deposit accounts, safety deposit boxes), and cash management services (for example, lock box or payroll processing). Other traditional lines of business for which banks have always earned fee income include trust services provided to a wealthy retail clientele and providing letters of credit (as opposed to immediate dispersal of loan funds) to corporate clients.

In recent years, advances in information, communications, and financial technologies have allowed banks to produce many of their traditional services more efficiently. These efficiencies not only reduced

TABLE 1

Selected sources of noninterest income at banking companies

Fee-generating activities: Nontraditional		
Investment banking		
Securities brokerage		
Insurance activities		
Merchant banking[a]		
Fee-generating activities: Traditional		
Traditional production methods	***New production methods***	
Deposit account services (e.g., safe-keeping, checking)	Deposit account services (e.g., online bill-pay, ATMs)	
Lending (e.g., letters of credit)	Lending (e.g., securitization, servicing)	
Cash management (e.g., payroll processing, traditional lock box)	Cash management (e.g., lock box check conversion to electronic ACH payments)	
Trust account services (e.g., wealth management)		

[a]A merchant bank invests its own capital in leveraged buyouts, corporate acquisitions, and other structured finance transactions. The merchant bank typically arranges credit financing, but does not hold the loans to maturity.
Source: Fitch (2000).

per unit costs, enhanced service quality, and increased customer convenience, but also represented a source of increased fee income for banks. Some examples are displayed in the third column of table 1. Advances in credit-scoring models and asset-backed securities markets have transformed the production of consumer credit and home mortgages from a traditional portfolio lending process, where banks earn mostly interest income, to a transaction lending process, in which banks earn mostly noninterest income (for example, loan origination fees and loan servicing fees). Advances in communications and information technologies have led to new production processes for transactions and liquidity services, such as ATMs (automated teller machines) and online bill-pay, and deposit customers have been willing to pay fees for these conveniences. (The phase out of Regulation Q ceilings on deposit interest rates assisted banks in this regard, allowing them to price depositor services in a more rational and competitive fashion.)

Similar to the noninterest income generated by nontraditional activities, the fee income derived by these new production methods is uneven across banking companies. Securitized lending processes generate significant scale economies, and as a result fee income from securitized consumer and mortgage lending has flowed predominantly (though not completely) to large banking companies. In contrast, the scaleable technologies necessary to produce ATM and Internet banking services are accessible to even relatively small banks.

Financial statement data

Taking advantage of the highly detailed financial statements that commercial banks and bank holding companies provide to their regulators, we collected data for established U.S. banking companies in 1986, 1990, 1995, 2000, and 2003. This multi-year, multi-company dataset allows us to observe how business strategies differ across banking companies in a given year and how banking strategies have changed over the past two decades as regulatory, technological, and competitive conditions have changed.[1] For the purpose of our analysis, an "established banking company" is either an independent commercial bank that is at least ten years old or a bank holding company (BHC) or financial holding company (FHC) that controls one or more commercial banks that are on average at least ten years old. These categories of banking companies are inclusive of all mature U.S. commercial bank charters and, as such, they include banking companies of all sizes—from small, independently organized community banks to large financial holding companies—that operate using a diverse array of banking business strategies.

We approach these data somewhat differently than most financial analyses of the commercial banking industry. First, we pay as much attention to bank income statements as we do to bank balance sheets. Financial analysis of commercial banks often concentrates on bank balance sheets, which display the most direct evidence of banks' traditional intermediation activities between depositors and borrowers. (Deposits are the largest single item on the liability side of most banks' balance sheets, and loans are the largest single item on the asset side of most banks' balance sheets.) But balance sheets have become an increasingly incomplete records of banks' profit-generating activities; they convey very little information about the fee-based activities that now generate over 40 percent of total operating income in the banking industry.

TABLE 2

Size of banking companies in DeYoung–Rice dataset, thousands of 2003 dollars, unless indicated otherwise

		1986	1990	1995	2000	2003
Number of banking companies		3,799	3,127	2,924	2,644	2,662
Assets	Mean	552,527	1,019,863	1,454,478	2,346,017	2,746,374
	Median	46,720	56,083	95,565	202,791	232,224
Operating income	Mean	25,701	53,062	78,535	142,446	157,582
	Median	2,085	2,431	4,663	9,362	10,536
Book value	Mean	33,387	62,097	115,193	182,256	225,723
	Median	4,358	5,252	10,406	18,230	21,475
No. of full-time employees	Mean	34.76	35.78	39.35	42.38	44.31
	Median	18	18	21	20	20
No. of branches	Mean	3.94	8.71	16.82	21.32	22.06
	Median	1	1	3	5	5

Some of these fee-based activities are traditional (like providing services to depositors and private banking clients); some are new to commercial banks (like investment banking, venture capital, and insurance underwriting); and some are traditional banking activities produced using new, nontraditional methods (like automated lending processes). Because income statements display the revenues and expenses generated by all of a bank's activities—whether or not they are represented on the balance sheet—we analyze bank income statements first before moving on to bank balance sheets.

Second, we construct financial ratios two different ways: We construct composite (or size-weighted) financial ratios using aggregate data for the entire commercial banking industry; and we construct bank-level (or unweighted) financial ratios using data from individual commercial banks. The composite financial ratios are informative about the overall product mix, financing mix, risk, and profitability of the commercial banking industry, but these ratios may not be descriptive of the "typical" commercial bank because large banks dominate the aggregate data. To the extent that a typical bank exists (and this is a problematic concept in itself, as discussed in the companion article that follows), it would be better described by taking the average of the bank-level ratios. For some financial ratios the size-weighted and unweighted averages have similar values; but for other ratios these two approaches yield substantially different values. As we shall see, these differences can reveal important information about how commercial banks make money. Large size allows banking companies to serve large corporate clients and provides them with access to low-cost, high-volume production, distribution, and marketing processes. But large size can make it difficult for banking companies to provide personalized retail service and/or build relationships with their small business loan customers.

The financial data for independent banks were drawn primarily from the Reports of Condition and Income (call reports), and the financial data for BHCs and FHCs were drawn primarily from the Federal Reserve Board FR Y-9C reports. These data were augmented with data from a number of other sources, including the Federal Reserve Board National Information Center's (NIC) structure database, the Federal Deposit Insurance Corporation's Summary of Deposits database, and the Center for Research on Stock Prices (CRSP). To be included in the dataset in any given year, a banking company had to be domestically owned, have positive amounts of loans and transaction deposits, have positive book value, and be FDIC-insured or own at least one commercial bank that was FDIC-insured. We express all data in thousands of 2003 dollars, unless otherwise indicated.

Table 2 displays some basic summary statistics for each of the years in our 1986–2003 sample period. The number of banking companies has declined over time for two reasons: nearly a thousand commercial banks failed during the first ten years of our sample period and, in each year of our sample period, hundreds of banking companies were merged or acquired. These trends were mitigated to some extent by the thousands of new banking companies that were started up during the 1980s and 1990s (entering our dataset upon turning ten-years old) and by the entry of some nonbank FHCs (investment banks, insurance companies, and securities firms) after 1999 under the provisions of the Gramm–Leach–Bliley Act. The size of the average banking company increased substantially during our sample period, in terms of assets, operating income, book value, employees, and branches.

Noninterest income: Evidence from the income statement

Table 3 displays income statement data from the five years contained in our 1986–2003 dataset. Each of the revenue, expense, and profit items is expressed as a percentage of operating income, except return on assets (ROA) and return on equity (ROE). The size-weighted ratios are indicative of the composition of total industry revenues, expenses, and profits. The unweighted ratio averages are indicative of the composition of revenues, expenses, and profits at the average bank.[2]

The most systematic change in bank income statements during our sample period is the increasing incidence of noninterest income, which now accounts for about 20 percent of operating income at the average commercial banking company (up from about 13 percent in 1986) and about 47 percent of total industry operating income (up from about 30 percent). In other words, today the banking industry generates slightly more than $1 of net interest income for every $1 of noninterest income, compared with just two decades ago when this industry multiple was well over $2. As discussed above, the increased importance of fee income at commercial banking companies is a direct result of structural changes like industry deregulation, new information technologies, and financial innovation. The companion article that follows discusses the implications of these changes for competitive strategies at commercial banking companies.

The expense data suggest that the banking industry has become more cost efficient over the past two decades—noninterest expenses currently consume about $0.59 of every $1 of operating income generated by commercial banking companies, down dramatically from about $0.69 in 1986. A large part of this decline is due to increased competitive pressure and the incentives this creates for banking companies to operate

TABLE 3

Income statement items, as a percent of operating income, except ROA and ROE

	1986	1990	1995	2000	2003
Number of banking companies	3,799	3,127	2,924	2,644	2,662
			Size-weighted averages		
Net interest income	70.1	65.2	64.1	51.2	52.9
Noninterest income	29.9	34.8	35.9	48.8	47.1
Noninterest expense	69.2	69.7	63.8	63.0	59.3
Labor expense	34.4	33.6	31.6	29.9	30.2
Full-time employees (workers per $mil.)	8.6	7.4	6.1	4.4	4.3
Premises expense	11.4	11.4	9.6	8.0	8.0
Other noninterest expense	23.4	24.6	22.5	25.1	21.1
Provisions for loan losses	14.6	18.1	4.7	7.6	7.3
Taxes and extraordinary items	1.2	2.9	10.9	10.6	9.9
Net income (ROS)	15.0	9.3	20.6	18.8	23.5
Return on assets (ROA)	0.0070	0.0048	0.0111	0.0114	0.0135
Return on equity (ROE)	0.1152	0.0789	0.1408	0.1471	0.1641
			Unweighted averages		
Net interest income	87.1	85.0	84.3	83.0	79.7
Noninterest income	12.9	15.0	15.7	17.0	20.3
Noninterest expense	67.4	69.5	65.7	64.6	66.2
Labor expense	34.5	35.4	34.6	35.0	36.7
Full time employees (workers per $mil.)	10.3	10.1	9.0	8.1	7.8
Premises expense	9.7	9.3	8.9	9.2	9.0
Other noninterest expense	23.2	24.8	22.1	20.4	20.4
Provisions for loan losses	18.1	8.3	3.4	5.2	4.9
Taxes and extraordinary items	18.0	13.9	12.4	13.4	11.2
Net income (ROS)	14.5	16.6	21.9	21.9	22.5
ROA	0.0066	0.0074	0.0106	0.0105	0.0105
ROE	0.0476	0.0682	0.1031	0.1064	0.1102

Note: ROS is return on sales.

more efficiently. Shifts in banking product mix and the introduction of new ways to produce and distribute traditional banking products likewise have contributed to this decline in expenses. Note that the number of full-time employees per dollar of operating income has fallen precipitously over time, while industry-wide labor expenses have declined only marginally and have actually increased at the average bank. These conflicting trends provide evidence that new banking products and production methods require a more highly skilled work force and, hence, higher salaries and benefits to attract and retain these workers. For example, while low-wage bank tellers have become less necessary due to ATMs and online banking, high-wage finance and information professionals have become more necessary to manage these systems and the growing array of products offered over them.

Labor expenses, premises expenses, and other noninterest expenses have all declined over time for large banks (which dominate the size-weighted ratio averages), but in contrast, only one of these three expense items has declined for the average bank (which is better represented by the unweighted ratio averages). As large banking companies have grown even larger via industry consolidation, they have increasingly benefited from scale economies that drive down per-unit costs; moreover, large banking companies are more likely to participate in high-volume, fee-based activities like automated lending, online banking, and mass marketing campaigns that benefit from scale economies. Naturally, the small banks cannot benefit as much from these economies of scale—not only because of their small size, but because many small banks practice more personal, relationship-based strategies that require relatively more customer-service labor inputs and relatively more physical spaces to interact with their customers.

Although expenses have declined less for the average banking company than for the industry overall, the proof of improved bank efficiency is in the profit pudding: net income has increased substantially, to just over 20 percent of operating income for both the average banking company and the industry as a whole. Because this return-on-sales (ROS) profit measure is a relatively uncommon way to express banking profitability, we also include the more familiar ROA and ROE measures, both of which have increased over time as well. This broad improvement in profitability has three fundamental causes: improved cost and revenue efficiency due to advances in information technology and financial processes; improved cost and revenue efficiency in response to the competitive pressures brought on by industry deregulation; and the generally improved banking environment starting in the mid-1990s, reflected in the table as reduced loan loss provisioning.

The ROA and ROE data suggest that the average bank—with an ROA of 1.05 percent and an ROE of 11.02 percent in 2003—is less profitable compared with the industry-wide aggregate ROA and ROE measures of 1.35 percent and 16.41 percent, respectively. As with many of the other differences we observe in the financial data, higher levels of noninterest activities at some banks also help explain these differences in the ROA and ROE measures. Because large banking companies tend to generate large amounts of fee income from activities that are not found on the balance sheet (for example, fee income from securitized lending activities), these banks will naturally appear to be more profitable using an ROA measure. Additionally, because large banking companies tend to be more diversified across product lines and customer bases and are more likely to use derivatives securities and complex modeling techniques to mitigate risk, they can operate with a smaller cushion of equity capital. Therefore, they will also appear to be more profitable using an ROE measure. Thus, for slightly different reasons, large banking companies will have higher traditional accounting performance measures than smaller banking companies, all else being equal.

Note, however, that ROS, ROA, and ROE are not risk-adjusted performance measures and, thus, using these measures to compare the profitability of different banks is an incomplete performance analysis. We compare and contrast risk-adjusted financial performance of different types of commercial banking companies in the companion article that follows this one.

Noninterest income: Evidence from the balance sheet

The dramatic increases in noninterest income over the past two decades have not occurred in isolation from other banking activities and, as such, they have left a trail not only on bank income statements but also on bank balance sheets. The increase in noninterest income has occurred in consort with changes in virtually every other area of commercial bank activities, including interest income, interest and noninterest expenses, bank asset mix, and bank funding sources. We now turn briefly to an analysis of bank balance sheets to illustrate these changes.

Assets

As displayed in table 4, there has been a marked change in asset mix since the mid-1980s. For the

average banking company (unweighted ratio averages), the big story is increased investment efficiency. In 1986 the average banking company had 50 percent of its assets invested in low-yielding assets like cash, securities, and fed funds, and only about 47 percent in loans. But by 2003, investments in loans at the average banking company were nearly twice as large as investments in lower yielding assets (61.1 percent versus 34.9 percent). These figures are clear indications that, despite the increased importance of noninterest income, the survival of the average banking franchise continues to depend on traditional intermediation from depositors to borrowers.

Low-yielding cash balances have also declined for the industry as a whole (size-weighted ratio averages), but investments in loans have also fallen substantially, from 62.3 percent to 52.5 percent of assets. But this does not necessarily indicate a reduction in investment efficiency. These data are consistent with a shift in the production functions of large banking companies away from traditional portfolio lending and its reliance on interest income and toward securitizable transaction lending (especially credit cards and home mortgages) that relies on noninterest income from loan origination and loan servicing fees. The 10 percentage point reduction in loan assets has been more than offset by a 12 percentage point increase in "other assets," such as the fair value of derivative instruments used to hedge against interest rate and foreign currency risk and receivables on the interest rate portion of asset-backed securities (IO strips).

Real estate loans have become a much more important part of bank loan portfolios over the past two decades. A number of factors played a role in this, including easier access to mortgage financing, the 1986 tax reform act that eliminated the consumer debt interest deduction but maintained the mortgage interest deduction, an increase in home ownership rates, the run-up in single-family home prices in many markets, as well as a need for banking companies to replace lost market share in commercial and industrial (C&I) loans. Between 1986 and 2003, C&I loans declined from 31.53 percent to just 18.90 percent of the overall industry loan portfolio, as large business borrowers began to bypass banks in favor of direct finance (for example, issuing commercial paper or high-yield debt), and nonbank competitors such as insurance companies and investment banks began to compete with banks for the remainder of the shrinking C&I loan market. For some banks, increased fee income from issuing letters of credit has softened the loss of C&I market share.[3] In contrast, both C&I loans and commercial real estate loans—on-balance sheet, relationship-based loans that generate interest income—have increased substantially for the typical banking company.

Financing and deposit mix

Deposits are the single most important source of financing for banking companies. As shown in table 5, about 57 percent of the banking industry's assets, and about 82 percent of the typical banking company's assets, were financed with deposits in 2003. Community banks use higher levels of deposit funding, and their noninterest income streams depend heavily on depositor service charges. However, even these high levels of deposit funding mark a decline over the past two decades, in favor of increased funding from federal funds, subordinated debt, "other liabilities," and equity financing. This reflects at least three developments: increased competition from nonbanks (for example, mutual funds, brokerage accounts) for household and business deposits; expanded ability of large banking companies to raise debt in financial markets (for example, commercial paper, subordinated debt); and regulations that now require banks to hold higher levels of equity capital than in the past.

The composition of deposits has also changed over time, and these trends reflect differences in the ways that large and small banks do business. For the typical banking company, transaction deposits have held relatively steady over the years—at about 28 percent of total deposits in general and about 15 percent of total deposits for banks' business clients (demand deposits). Thus, relationships with depositors and access to the payments system continue to be essential parts of most banking companies' business strategies. In contrast, transaction deposits have declined dramatically at the industry level (from 29.8 percent to 15.0 percent of total deposits) since 1986.

A closer look at noninterest income and risk

Both traditional and nontraditional banking activities generate noninterest income. Traditional fee-generating activities include transaction services for retail and business depositors (although in recent years a growing percentage of these fees has been charged for nontraditional technologies like online bill-pay) and fiduciary services for high net worth retail clients. Nontraditional fee-generating activities include investment banking, insurance underwriting and agency, and venture capital. Finally, banking firms generate a substantial amount of noninterest income by using nontraditional methods to produce traditional banking services. For example, in a traditional banking model, loan servicing fees and securitization fees do not exist, because banks hold the loans they originate in their own portfolios and service these loans themselves.

TABLE 4

Asset items, as a percent of total assets

	1986	1990	1995	2000	2003
Number of banking companies	3,799	3,127	2,924	2,644	2,662
			Size-weighted averages		
Cash	11.5	8.7	6.5	5.1	4.5
Securities	16.5	16.6	18.1	16.2	17.6
Fed funds sold	3.1	2.4	3.9	5.4	6.5
Loans	62.3	64.3	58.8	57.0	52.7
Allowance for loan losses	(0.9)	(1.6)	(1.2)	(1.0)	(0.9)
Fixed assets	1.6	1.7	1.6	1.2	1.1
Other assets	5.9	8.0	12.4	15.9	18.4
Loan items as % of loans					
Real estate loans	32.35	40.02	43.37	45.04	53.74
Residential mortgages	N/A	20.37	26.29	25.97	32.50
Home equity loans	N/A	3.00	3.30	3.51	6.96
Commercial real estate loans	N/A	17.45	15.19	17.33	19.47
Agricultural land loans	N/A	0.57	0.64	0.69	0.73
Consumer loans	21.25	19.41	13.81	11.08	17.09
Credit cards	N/A	0.18	7.24	5.70	6.96
Commercial and industrial loans	31.53	28.62	25.42	27.50	18.90
Agricultural production loans	1.39	1.09	1.07	0.94	0.79
Other loans	N/A	10.87	16.33	15.44	9.48
			Unweighted averages		
Cash	9.1	7.2	5.5	4.8	5.2
Securities	33.0	32.9	32.1	26.3	26.1
Fed funds sold	7.9	6.5	5.1	3.6	3.4
Loans	46.6	49.7	54.0	61.7	61.1
Allowance for loan losses	(0.7)	(0.9)	(0.9)	(0.9)	(0.9)
Fixed assets	1.5	1.5	1.6	1.8	1.8
Other assets	2.5	2.8	2.4	2.6	3.2
Loan items as % of loans					
Real estate loans	40.53	46.77	55.39	60.39	65.91
Residential mortgages	N/A	27.03	29.74	28.53	27.07
Home equity loans	N/A	1.03	1.46	1.72	2.62
Commercial real estate loans	N/A	13.90	20.50	26.89	33.51
Agricultural land loans	N/A	5.61	4.90	4.72	5.04
Consumer loans	23.89	20.77	16.95	13.23	10.46
Credit cards	N/A	0.25	0.71	0.49	0.44
Commercial and industrial loans	3.72	6.09	8.27	11.73	10.44
Agricultural production loans	15.05	13.20	9.59	6.90	5.94
Other loans	N/A	12.76	9.34	7.21	6.72

Notes: Columns may not sum to 100 percent due to rounding errors. N/A indicates that data were not available in 1986. The "other assets" category combines a variety of assets that are not separately reported to regulators at the banking holding company level, including (but not limited to) interest receivable on loans and securities, derivative securities not held for trading purposes, equity securities without readily determinable fair values (for example, stock in a Federal Home Loan Bank or equity holdings in corporate joint ventures), prepaid expenses, repossessed property such as automobiles and boats, credit or debit card sales slips in the process of collection, and assets held in charitable trusts.

By some measures, noninterest income might be characterized as a large-bank phenomenon. As shown in table 6 (p. 47), noninterest income accounts for only about $1 in $5 of operating income at the average banking company with assets less than $1 billion but about $1 in $2 of operating income at the average banking company with assets greater than $25 billion. Moreover, the lion's share of noninterest income is being generated by a very small number of banking companies: In our sample, 84 percent of all noninterest income in 2003 was generated by just 1 percent of the banking companies (not shown).

Scale economies in production are one reason that noninterest income represents such disparate amounts

TABLE 5

Liability and equity items, as a percent of total assets

	1986	1990	1995	2000	2003
Number of banking companies	3,799	3,127	2,924	2,644	2,662
			Size-weighted averages		
Deposits	74.1	75.7	66.7	58.2	57.2
Fed funds purchased	7.6	5.2	7.2	8.7	9.3
Subordinated debt	0.5	0.7	1.7	1.8	1.9
Other liabilities	11.7	12.3	16.5	23.5	23.4
Equity	6.0	6.1	7.9	7.8	8.2
Deposit items as % of deposits					
Transactions deposits	29.8	25.8	27.0	17.4	15.0
Demand deposits	21.9	17.6	18.9	13.4	10.5
Nontransactions deposits	70.2	74.2	73.0	82.6	85.0
Savings and MMDAs	38.0	34.1	42.3	51.2	61.4
Small CDs	20.0	26.8	23.1	18.9	12.7
Large CDs	12.2	13.3	7.8	12.5	11.0
			Unweighted averages		
Deposits	88.3	87.8	85.6	82.7	81.8
Fed funds purchased	0.7	0.8	1.1	1.8	1.7
Subordinated debt	0.1	0.1	0.1	0.1	0.1
Other liabilities	1.4	1.8	2.3	5.0	6.0
Equity	9.4	9.5	10.7	10.4	10.3
Deposit items as % of deposits					
Transactions deposits	28.7	27.4	30.4	27.0	27.6
Demand deposits	16.5	14.5	16.4	15.3	15.2
Nontransactions deposits	71.3	72.6	69.6	73.0	72.4
Savings and MMDAs	23.9	20.0	23.2	23.7	29.5
Small CDs	38.0	41.7	35.6	33.4	27.6
Large CDs	9.0	10.6	10.5	15.3	14.8

Notes: Columns may not sum to 100 percent due to rounding errors. The "other liabilities" category combines a variety of liabilities that are not separately reported to regulators at the banking holding company level, including (but not limited to) accounts payable, deferred compensation, trust preferred securities, dividends declared but not yet payable, allowances for credit losses on off-balance-sheet credit exposures, and selected insurance subsidiary liabilities (such as unearned premiums and claims expense reserves). MMDA is money market deposit account and CD is certificate of deposit.

of income at large and small banking companies. For example, loan servicing and other automated techniques, which generate large amounts of fee income relative to traditional production techniques, are most cost-effective when used at high volumes. Similarly, investment banking and other nontraditional banking products that generate large amounts of fee income tend to be practiced at banking companies that are large enough to service big corporate clients.

The composition of noninterest income also differs across banking companies of different sizes. Large banking companies generate disproportionately more noninterest income from securitizing and servicing mortgage and credit card loans, because the automated production processes used to produce these services exhibit substantial scale economies. Similarly, large banking companies are better able to employ the concentrations of financial experts and develop the institutional information databases necessary for the production of investment banking, insurance underwriting, and private banking (fiduciary) services. However, there are other areas in which smaller banking companies generate a higher percentage of noninterest income than larger banking companies. Because small banking companies rely more on core deposit funding (such as household and small business checking accounts) than do larger banks, deposit service charges comprise a large part of their fee income base. And fee income from the sale of insurance products shows no size bias—possibly because small banking companies have been successful at cross-selling insurance products to their existing household and small business clients.

Noninterest income and financial performance

As discussed above, increased competition from nonbanks and out-of-state banks in the years following deregulation put downward pressure on the profitability of commercial banking companies. Banks were sensitive to these coming challenges long before deregulation was fully implemented. For example, in a 1983 study the Texas Bankers Association concluded that if the banking industry was "to remain profitable in a deregulated environment, it must fundamentally change the way in which it makes money." The study found that in order to offset expected declines in net interest margins due to post-deregulation competition, fee income would have to increase from about 15 percent of total income to over 50 percent of total income by 1986 (Lane, Friars, and Goldberg, 1983). As we have seen above, many banking companies heeded the spirit of this advice by offering an expanded menu of fee-based products and services.

The financial consequences of these strategic shifts toward fee income were not well understood at the time. The upward trend in noninterest income during the 1990s was generally believed to have two risk-reducing effects: shifting banks' income mix away from intermediation-based activities would reduce banks' exposure to credit risk and interest rate risk, and shifting banks' income mix toward fee-based financial products and services would reduce earnings volatility via diversification effects. As late as 2000, many bankers continued to believe that fee income would be a stable income stream: "Indeed, shareholders and analysts alike have grown fond of the earnings, diversity, growth potential, and market insulation that fees provide" (Engen, 2000).

To some extent industry-wide trends in income mix and profitability offered superficial support for this view. As shown in figure 2, between 1980 and 2003 noninterest income doubled as a percent of total industry operating income at U.S. commercial banking companies, while at the same time aggregate industry profitability (ROE) not only increased but became more stable. But such an analysis is indeed superficial and it ignores growing evidence to the contrary. Recent empirical studies indicate that although an increase in noninterest income may improve bank earnings, this seldom occurs without concomitant changes in interest income, variable inputs, fixed inputs, financing structure, and other changes that have risky implications for the variability of bank earnings.

DeYoung and Roland (2001) suggest three reasons that noninterest income may increase the volatility of bank earnings. First, loans that are held in a bank's portfolio—especially loans to businesses—are relationship based. That is, banks have close ties to their borrowers that allow them to ascertain borrower creditworthiness by building up a storehouse of private information about the borrower and to monitor the activities of the borrower going forward. Because these informational ties are costly for both the bank and the borrower to replicate, relationship-based loans often have high switching costs. In contrast, some fee-based activities are not relationship based and, hence, have low switching costs, such as fees from originating a mortgage or from non-customers using a bank's ATM machines.[4] Thus, despite exposing the bank to credit risk and fluctuations in interest rates, interest income from loans may be less volatile than noninterest income from many fee-based activities.

Second, a bank that shifts its product mix from traditional asset-based, interest-generating activities to nontraditional fee-based activities tends to increase its "degree of operating leverage." For example, within the context of an ongoing lending relationship, the main input needed to produce more loans is a variable input (that is, interest expense); in contrast, the main input needed to produce more fee-based products is typically a fixed or quasi-fixed input (that is, labor expense). Thus, fee-based activities may require greater operating leverage than lending activities, which makes bank earnings more vulnerable to declines in

FIGURE 2

"Is there a connection between high levels of noninterest income and earnings stability?"

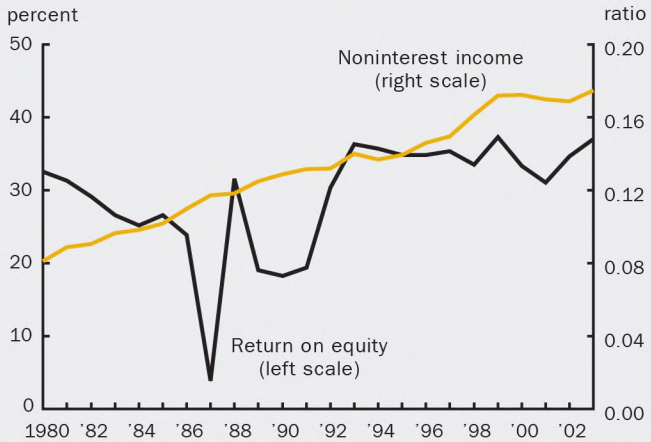
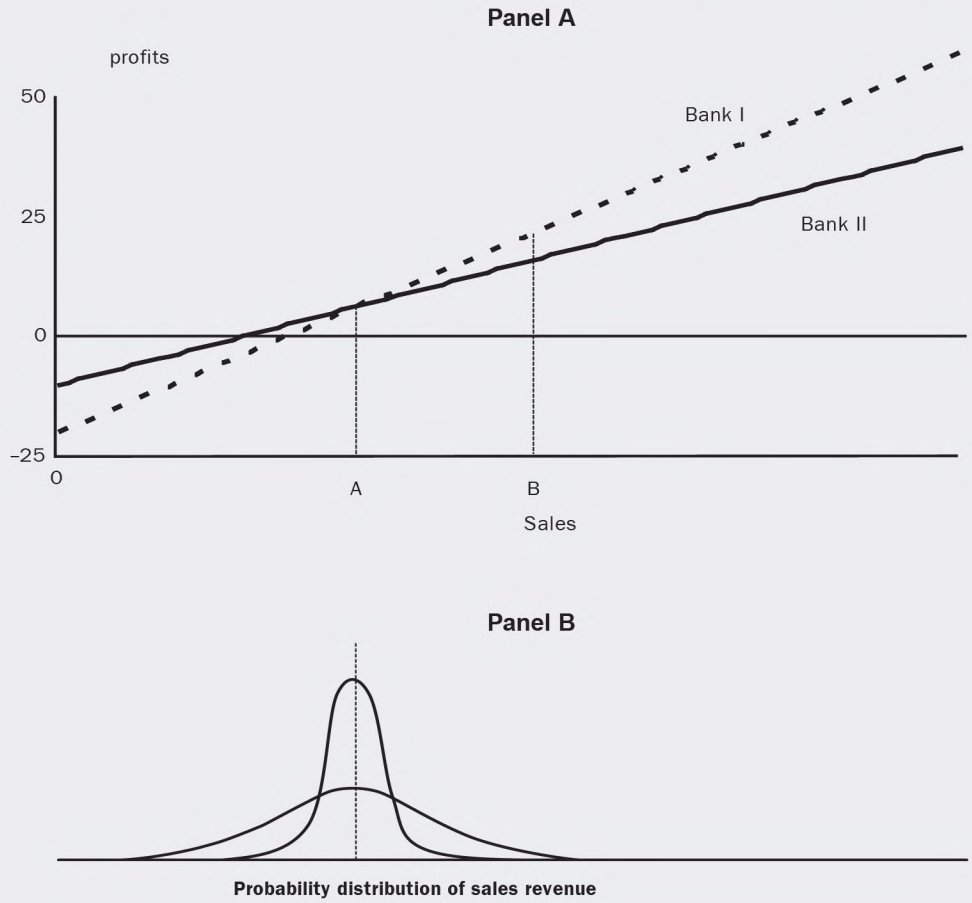


FIGURE 3

Bank I has high fixed and low variable costs (high degree of total leverage); Bank II has low fixed and high variable costs (low degree of total leverage)



Source: Modified from DeYoung and Roland (2001).

revenues. Third, most fee-based activities require banks to hold little or no fixed assets, so unlike interest-based activities like portfolio lending, fee-based activities like trust services, mutual fund sales, and cash management require little or no regulatory capital. This allows banks to finance a greater amount of their income-generating activities with debt, which increases fixed interest expenses. In other words, fee-based activities allow banks to use a greater "degree of financial leverage" than lending activities.

The theoretical implications of these phenomena are illustrated in figure 3. Banks with low fixed costs and high variable costs, that is, a low degree of total leverage, are represented in panel A by the solid line (labeled Bank II). Banks with high fixed costs and low variable costs, that is, a high degree of total leverage, are represented by the dashed line (Bank I). Theoretically, banks with large amounts of fee-based income will have a high degree of total leverage like Bank I, so that any given change in revenue (along the horizontal axis) gets amplified into a greater change in earnings (along the vertical axis). For example, when sales equals amount A, both Bank I and Bank II are earning about 5 units of profit. If sales revenue for both firms increases to B, Bank I will realize a much greater change in profits, from 5 units to 25 units, while the earnings for Bank II will increase from 5 units to about 15 units.

To make matters worse, the distribution of sales revenues can also vary across banks. Panel B illustrates two probability distributions for sales revenue. The flatter of the two distributions represents sales that are more volatile, while the more bell-shaped of the two represents sales that are less volatile, that is, they will not vary as much around the mean (represented

by the dotted line.). Theoretically, revenues derived from non-relationship-based fee-generating activities with low switching costs (that is, home mortgage origination) or high sensitivity to the business cycle (that is, investment banking) can be more volatile than revenues from traditional relationship-based banking lines of business. A bank with more of these types of activities would be likely to have sales revenues represented by the flatter of the two distributions. This would compound the change in profits illustrated in panel A because sales at Bank I would be more likely to change from quarter to quarter, all else being equal.

Using data from U.S. banks during the 1990s, DeYoung and Roland (2001) find evidence consistent with these theoretical conjectures. They demonstrated that income from fee-based activities was more volatile than income related to traditional intermediation activities (that is, interest from loans, interest from securities, and service charges from deposits), and that the degree of total leverage tends to be greater at banks that generate large amounts of fee income from nontraditional activities. They find also that the type of fee-based activity in which the bank engages makes a difference. Increased reliance on trading activities increases the volatility of overall bank revenues, while increased reliance on charging fees to depositors reduces the volatility of overall bank revenues.

Studies that followed generated similar findings. Stiroh (2004a) finds that an increased focus on noninterest income is associated with declines in risk-adjusted performance. In another study, Stiroh (2004b) finds potential diversification benefits for banks that offer a variety of different fee-based services, but no diversification benefits for banks that produce a combination of interest-based and fee-based services. In DeYoung and Rice (2004), we find that increases in noninterest income are associated with higher, but more volatile, accounting rates of return, resulting in reduced risk-adjusted returns.

We also find that an increase in noninterest income does not necessarily indicate that intermediation activities have become less important at banking companies. If intermediation activities have become less important for banks over time, we argue in DeYoung and Rice (2004), then it stands to reason that the correlation between bank profitability and bank net interest margin would grow weaker over time. But to the contrary, we show that the correlation of bank earnings (ROE and ROA) and net interest margin has not grown weaker over time and may have actually strengthened slightly. Finally, we find that well-managed banks tend to focus on a narrow set of fee-based activities, most of which are unrelated to either traditional core deposit business or trading activities. This includes (among other items) fees from the sale of mutual funds and insurance policies, fees from securitization activities, income from loan servicing, fees from providing trust services, and income from providing cash management services. In DeYoung and Rice (2004), we conclude that increased noninterest income is co-existing with, rather than replacing, intermediation activities at the typical commercial bank and that traditional intermediation activities remain the core activity of most profitable banks.[5]

A closer look at noninterest income and payment activities

Providing payment services is an under-appreciated source of noninterest income for banking companies and its importance may be growing. Payment-related information technology at banks was expected to grow about 37 percent between 2003 and 2004 (Access Intelligence, LLC, 2004). Wachovia Corp. reorganized its payment operations to create a centralized payment division (Wade, 2003), while Bank of America Corp. overhauled its internet banking and bill-pay website (Bills, 2003). As the menu of payment products and services available to consumers increases, bankers have recently acknowledged that "only within the last two to three years has there been a realization of the importance of payments" (Wade, 2003).

Since customers generally use transaction accounts to make and receive payments, banking companies play a natural role in the payments system.[6] And although competition from nonbanks using new payment technologies has increased, it is likely that banks will remain primary providers of payment products and services, because banks have two unique features with regard to the payments system that nonbank firms do not share.

First, financial institutions have the ability to offer settlement activities. Settlement here is defined as the irrevocable transfer of funds between parties in a payments system.[7] While nonbank firms are very much involved in the processing of payments in the U.S. economy, only financial institutions can settle payment transactions, because all noncash payment transactions, except for on-us transactions, require the transfer of funds between two financial institutions. For example, Fiserv is a large vendor, or "third party provider," of transaction services, such as customer account processing and check processing and imaging, but it does not settle checks with customers' accounts. Only banks can settle their customers' accounts.

Second, because the payments system is heavily reliant upon deposit-based instruments, banks are in a unique position to profit by cross-selling payment-based, payment-related, and non-payment-related products and services to their deposit customers. For example, banks offer customers a broad menu of payment methods with which to access the funds in their deposit accounts—such as checks, debit cards, direct debit for paying bills, direct deposit for receiving paychecks, and online bill paying—and offer payment products peripheral to customers' accounts, such as credit cards and home equity lines of credit.

Depending on their business model and competitive strategy, banks can and do charge fees for these payment-related services. Chakravorti and Kobor (2003) suggest that banking companies use two different approaches for using payment activities to enhance their profits: a stand-alone product approach and a product-bundling approach. Stand-alone payment strategies are highly specialized; some examples include securities processing and handling, management of large personal and corporate trust accounts, and correspondent banking services. These lines of business tend to generate revenue streams that are independent from banks' other activities and tend to use specialized (and often large scale) production processes as well.

In contrast, banks that use a product-bundling strategy will market and price their payment products in conjunction with other retail or wholesale products. Retail products would include, for example, services tied to personal deposit accounts, while wholesale products would include corporate transactions. Although payment products may not contribute directly to profits in this approach, including them in a bundle of related services can increase the retention of deposit customers. In contrast to fee-based activities that are not relationship-based (such as home mortgage origination, as discussed above), products and services linked to deposit accounts can be "sticky," that is, they make it less convenient for deposit customers to transfer their funds to another institution. For example, deposit customers that use direct deposit and automatic bill payments are less likely to switch to another institution, because they will have to incur the costs (time and effort) of undoing these automated arrangements at their current bank and setting them up again at their new bank. Banking companies have become increasingly aware that their profits can be enhanced by offering costly new relationship-based services (such as expanded networks of branches or ATMs) at low prices or for a fee (where payment is in the form of foregone interest), because the switching costs associated with these services can embed the customer more firmly in the long run (Furst, Lang, and Nolle, 1998; Kiser, 2002; DeYoung and Hunter, 2003).

Given that banks use payments for vastly different strategic reasons, it is difficult to model and measure how payments contribute to profits. Two recent Federal Reserve studies have grappled with this problem. Radecki (1999) estimated that the top 25 U.S. banking companies derived between one-third and two-fifths of their operating revenue (noninterest income plus net interest income minus provisions for loan losses) from payment-related activities in 1996. This innovative study was the first to estimate the proportion of income at banks that is attributable to payment activities, and it made two essential contributions to our knowledge of the role of payment services in banking strategy. First, by expanding the definition of payment activities to include "transaction services performed outside a deposit account relationship," Radecki revealed the surprising depth of payment activities at large banking companies. Second, his study reminds us that payment services are integral to the strategic and financial functioning of most banking companies, because they are intertwined with the production of depositor-taking and information-intensive lending activities.

Rice and Stanton (2003) updated and expanded Radecki's study by estimating the volume of payment-driven revenues at the top 40 bank holding companies in 2001. In order to obtain a larger sample of banking companies, Rice and Stanton drew their data from the financial reports (call report and Y9) that U.S. banking companies file with their regulators, rather than relying on banking company annual reports, which do not offer consistent information on payment-driven revenues in several important categories. In addition, these authors adjusted some of Radecki's approximation methods, which may have over-allocated some bank revenues to payment activities. They conclude that payment revenue accounts for 16 percent to 19 percent of operating revenue—a substantially lower estimate than Radecki's but still a surprisingly large contribution to the overall revenue streams of banking companies.

Estimating the importance of payment revenues

To estimate the contribution of payment activities to the income of the 2,662 banking companies in our sample in 2003, we apply the estimation employed by Rice and Stanton, using the call and Y9 reports. (Most of these data are not available prior to 2001.) The method identifies four separate sources of payment-driven revenues:

TABLE 6

Noninterest income items, unweighted averages

	All banking companies	Assets < $1 billion	Assets > $1 billion–< $25 billion	Assets > $25 billion
		% of total operating income		
Noninterest income	**20.4**	19.0	30.1	49.2
		% of total noninterest income		
Fiduciary activities	**4.2**	3.5	9.7	15.0
Deposit fees	**51.6**	53.8	35.2	17.5
Loan servicing fees	**1.8**	1.8	1.9	4.1
Investment banking	**2.1**	1.7	4.9	13.5
Venture capital	**–0.01**	–0.01	–0.04	0.27
Securitization fees	**0.2**	0.1	0.9	10.2
Insurance agency	**3.4**	3.3	3.5	2.8
Insurance underwriting	**0.1**	0.1	0.4	1.9
Gains on asset sales	**9.5**	9.2	12.0	6.5
Other noninterest income	**27.1**	26.5	31.7	28.3

Notes: Columns may not sum to 100 percent due to rounding errors. The "other noninterest income" category combines fee income from a variety of traditional and nontraditional banking activities that are not separately reported to regulators at the banking holding company level, including (but not limited to) fees for retail services such as mutual fund sales, safety deposit boxes, and credit cards, and fees for commercial services such as cash management, standby letters of credit, loan commitments, correspondent banking services, and financial consulting.

Traditional service charges on deposit accounts

Traditional service charges on deposit accounts are composed of two parts: the explicit fees charged to depositors (displayed in table 6) and the foregone interest revenue implicitly charged to depositors. It is easy to overlook the fact that depositors compensate banking companies for the convenience of having transaction accounts by foregoing interest on their account balances. Customers earn no interest on demand deposit account (DDAs) and earn below-market rates on deposits in negotiable order of withdrawal (NOW), savings, and money market accounts (MMDAs). These low-cost funds are key to banking companies' traditional intermediation activities, in which they earn profits by reinvesting these funds in higher yielding (and higher risk) market-rate loans and investments.

Foregone interest revenue is relatively straightforward to estimate. For our calculations we assume that deposits in all accounts earn the banking company the federal funds rate (that is, the bank's alternative funding rate). For each type of deposit account, we take the average spread between the federal funds rate and the deposit rate and multiply it by the aggregate balance in each type of deposit account. We then sum up the interest foregone by depositors by account type: DDAs, NOW and other interest checking accounts, and MMDA and other savings accounts. These explicit (service charges) and implicit (foregone interest) charges on deposit accounts make up the lion's share of payment-driven revenue at the typical banking company. Radecki (1999) contains an in-depth discussion on estimating foregone interest payments from banking companies.

Trust and investment services income

Some portion of banks' income from fiduciary (trust) activities is payment-related; as such, this should be included in the aggregate estimates of payment-driven revenues. Estimating the amount of payment-related trust revenues, however, is extremely difficult. Depending on the type of trust account that is managed or held by a BHC's trust department, the BHC will earn a wide range of revenues from payment activities. At one end of the spectrum are trust accounts, where no cash will be distributed nor payments made to the customer in the foreseeable future. A personal trust containing non-dividend-paying market securities is a good example—the main fees charged to that trust by the BHC are portfolio management fees, not payment-activity fees. At the other end of the spectrum are trusts that pay out monthly distributions of income to the beneficiary of a trust account that does not require much, if any, portfolio management. The majority of the activity in this type of account is payment related, so the majority of revenue earned by the BHC can be attributed to payment activities.

Because the available data do not allow us to observe these differences in the intensity of fee income in the trust accounts of different banking companies, we create both a high estimate and a low estimate of these revenues for each bank. At the low end of the

spectrum we only include revenues generated from "custody and safekeeping accounts," since most of the revenue earned in these accounts is derived directly from payment-related activities. We acknowledge that this excludes payment-related activities in other types of accounts, such as retirement and corporate accounts; however, some of these revenues are captured above as service charges and foregone interest revenues associated with deposit accounts. At the high end of the spectrum, we add in revenues from employee benefit accounts and corporate trust and agency accounts. This may overstate payment-related revenues, because the revenues recorded in these additional categories include fees from some nonpayment-related activities.

Credit cards

Payment-related credit card fees are also difficult to measure. Fee income from credit cards includes late payments, interest on credit card balances above the cost of a traditional loan, finance charges for cash advances, fees for handling transactions on behalf of merchants and card holders, and interchange fees for credit card purchases. Not all of these fees can be considered payment-related, however, and the situation is complicated further by the securitization of credit card assets, which moves a considerable amount of this activity off of the main balance sheet of the banking company. The banking company continues to earn a portion of the revenue from securitized credit card receivables through the excess collateral from annual fees and other payment-related service charges, and so we must account for revenue generated by those assets as well.

Following Rice and Stanton (2003), we estimate the payment-related revenue from both the on-balance-sheet and off-balance-sheet credit card receivables. Based on data from Visa and MasterCard (Credit Card Management, 2001), Rice and Stanton estimate that about 17 percent of all credit card revenues are derived from payment services. Hence, we estimate both on-balance-sheet and off-balance-sheet payment-related credit card revenue by multiplying total revenue associated with credit cards by 0.17.

ATM revenues

Beginning in 2001, the call and Y9 reports require that all banks report the "income and fees from automated teller machines" (ATMs) in the category of "other noninterest income" when it exceeds 1 percent of gross interest income plus noninterest income. As a result, some banking companies show ATM revenues of zero (or missing), when in fact they are just below the 1 percent threshold. This will obviously result in an understatement of ATM revenues in our estimates.

Estimation results

Table 7 displays our estimates of payment-driven revenue for our sample of 2,662 banking companies in 2003, expressed as a percentage of operating income and broken out by bank size and payment category. We find that payment-related income comprises about 21 percent of operating revenue.[8] Given that we arrive at this estimation using a modified version of the Rice and Stanton (2003) methodology, it is not surprising that this figure is more in line with their previous estimates than with the larger estimates made by Radecki (1999). Still, our estimated figure of 21 percent remains substantial; it is nearly twice as large as the 12 percent of operating income that the average bank generates from fee-based activities that are unrelated to payments.

The importance and the mix of payment-related fees vary considerably by banking company size. For small banks, payment-related income is twice as large as non-payment-related fee income (21.12 percent versus 10.45 percent), while these figures are reversed for large banking companies (17.96 percent versus 38.55 percent). The average small banking company earns 97 percent of its payment-related revenue through traditional deposit accounts (deposit fees plus foregone interest), compared with only about 65 percent for the average large banking company. The typical large bank earns the balance of its payment-related revenue primarily through fiduciary income on trust accounts (about 23 percent) and credit card fees (about 10 percent).

Some large banking companies specialize in payment activities. State Street Corporation, for example, is the largest servicer of mutual funds and pension plans in the U.S., with more than $9 trillion in assets under custody. It earns an estimated 36 percent of its operating revenue from fiduciary fees on those custodial accounts. Other large banking companies emphasize payment activities from more traditional sources—for example, as mentioned above, Bank of America reorganized its online bill payment webpage, eliminated its monthly fee, and saw its usage more than double between 2002 and 2003. As a result, the bank achieved both "higher deposits and the higher retention benefits" (Bills, 2003).

Conclusions and Implications

Clearly, banking companies do more than just intermediate between depositors and borrowers. The industry has never limited itself to simply earning interest margins, and over time it has moved further away from that stylized version of "how banks make money." The most telling symptom of this movement is the remarkable increase in noninterest income at

TABLE 7

Payment-related income, unweighted averages

	All banking companies	Assets < $1 billion	Assets > $1 billion–< $25 billion	Assets > $25 billion
		% of total operating income[a]		
Income from payments-related activities (estimates)	**20.68**	21.14	17.06	17.96
Noninterest income unrelated to payment activities	**11.76**	10.45	20.86	38.55
		% of total payment-related income		
Deposit fees	**56.16**	56.32	55.54	44.04
Foregone interest revenue	**40.02**	40.75	33.98	20.55
ATM fees[b]	**1.93**	1.94	2.11	2.06
Fiduciary fees[c]	**1.16**	0.62	5.27	23.44
Credit card fees	**0.72**	0.38	3.10	9.91

[a]For the analysis in this table, operating revenue is defined net of provisions for loan and lease losses to remain consistent with the previous literature on this topic.
[b]Banking companies report ATM (automatic teller machine) fees only when they are at least 1 percent of total income.
[c]Based on the "upper bound" estimates of trust revenues.
Note: Columns may not sum to 100 percent due to rounding errors.

commercial banks, which by some measures now accounts for nearly half of the industry's income. For some banks, increases in noninterest income flow from new lines of business—such as investment banking, securities brokerage, and insurance agency—that were made possible by a historic dismantling of restrictive financial regulations in the 1990s. For other banks, increases in noninterest income flow from producing traditional banking services—such as securitized lending and electronic payment services—with new production processes that were made possible by advances in information technology, communications channels, and financial processes. Many banks have done both.

These changes have some surprising implications for the performance of financial institutions. Conventional industry wisdom held that rebalancing bank income away from interest income and toward noninterest activities and fee income would make banking companies less risky. Replacing interest income—with its sensitivity to unpredictable movements in interest rates and the business cycle—would reduce the volatility of bank income and expanding into nontraditional fee-based activities would yield risk-reducing benefits of diversification. However, recent research suggests that, at least so far, this has not come to pass. Diversification gains from fee-based activities appear to be scarce, and although there is some evidence that fee income can pump up bank earnings, this also tends to make bank earnings more, not less, volatile.

Perhaps just as surprising is the realization that traditional banking activities, namely, the provision of payment services, generate the lion's share of noninterest income at most banking companies. We extend the work of Radecki (1999) and Rice and Stanton (2003) and find that for the average banking company with assets less than $1 billion, payment-related revenues account for about 20 percent of total operating income and about two-thirds of total noninterest income.

Of course, the concept of "the average banking company" has become less meaningful over time, because technological change and industry deregulation have permitted (and the resulting increase in competition has encouraged) banking companies to experiment with innovative products, production processes, organizational forms, and business strategies. We explore the implications of these developments in the companion article that follows, which compares the financial performance of banking companies across different business strategies—from community banking to private banking to corporate banking and beyond.

NOTES

[1]We start our empirical analysis in 1986 because this is the first year that detailed financial data are available for bank holding companies.

[2]We truncated the values of the bank-level ratios at the 1st and 99th percentiles of their sample distributions (but did not discard those observations) in order to reduce the influence of outlying values.

[3]Fees from letters of credit (not shown) increased fourfold at the average banking company between 1986 and 2003, from approximately $2.40 to approximately $8.60 for every $1,000 of loans on the balance sheet.

[4]See Knittel and Stango (2004) for a discussion of ATM surcharges.

[5]This finding is reminiscent of the arguments made by Boyd and Gertler (1994) a decade earlier in a paper titled "Are banks dead? Or are the reports greatly exaggerated?"

[6]The payments system consists of a legal framework, rules, institutions, and technical mechanisms for the transfer of money. As such, it is an integral part of the monetary and financial system in a smoothly operating market economy (Hancock and Humphrey, 1997).

[7]For a thorough review of settlement and clearing systems, see the Committee on Payment and Settlement Systems (CPSS) Resources located on the Bank for International Settlements (BIS) website at http://www.bis.org/cpss/index.htm.

[8]To remain consistent with the previous literature on this topic, we define operating revenue net of loan loss provisions in this part of our analysis. Thus, payment-related revenue is calculated as the sum of the four payment-related components divided by (noninterest income plus net interest income minus loan loss provisions). Since foregone interest revenue is not included on the balance sheet (banks are not required to account for foregone interest), payment-related income plus noninterest income unrelated to payments will not add up to total noninterest income.

REFERENCES

Access Intelligence, LLC, 2004, "Why banks are stepping up payments-related IT spending," *Item Processing Report*, August 12.

Bills, Steve, 2003, "B of A upgrades payments section of its website," *American Banker*, June 20.

Boyd, John H., and Mark Gertler, 1994, "Are banks dead? Or are the reports greatly exaggerated?," in *Proceedings from a Conference on Bank Structure and Competition*, Chicago: Federal Reserve Bank of Chicago.

Chakravorti, Sujit, and Emery Kobor, 2003, "Why invest in payment innovations?," Federal Reserve Bank of Chicago, Emerging Payments, occasional paper series, No. 2003-1B.

Credit Card Management, 2001, *Card Industry Directory*, New York: Faulkner & Grey Publishers.

DeYoung, Robert, and William C. Hunter, 2003, "Deregulation, the Internet, and the competitive viability of large banks and community banks," in *The Future of Banking*, Benton Gup (ed.), Westport, CT: Quorom Books, pp. 173–202.

DeYoung, Robert, and Tara Rice, 2004, "Noninterest income and financial performance at U.S. commercial banks," *Financial Review*, Vol. 39, No. 1, pp. 101–127.

DeYoung, Robert, and Karin P. Roland, 2001, "Product mix and earnings volatility at commercial banks: Evidence from a degree of leverage model," *Journal of Financial Intermediation*, Vol. 10, pp. 54–84.

Engen, John R., 2000, "Fee income: Carving out a strategy," *Bank Director Magazine*, Quarter 3.

Fitch, Thomas P., 2000, *Dictionary of Banking Terms*, Hauppauge, NY: Barron's Educational Series, Inc.

Furst, Karen, William Lang, and Daniel Nolle, 1998, "Technological innovation in banking and payments: Industry trends and implications for banks," *Quarterly Journal*, Office of the Comptroller of the Currency, Vol. 17, No. 3, pp. 23–31.

Hancock, Diana, and David B. Humphrey, 1997, "Payment transactions, instruments and systems: A survey," *Journal of Banking & Finance*, Vol. 21, pp. 1573–1624.

Kiser, Elizabeth K., 2002, "Household switching behavior at depository institutions: Evidence from survey data," Federal Reserve Board, Financial and Economics Discussion Paper Series, working paper, No. 2002-44.

Knittel, Christopher R., and Victor Stango, 2004, "Incompatibility and consumer demand: Evidence from ATMs," draft.

Lane, Robert B., Eileen M. Friars, and Ellen S. Goldberg, 1983, "Banks had better change the way they make money; If they don't margin declines under deregulation may even turn earnings negative," *American Banker*, October 11.

Radecki, Lawrence, 1999, "Banks' payments-driven revenues," *Economic Policy Review*, Federal Reserve Bank of New York, Vol. 4, No. 2, July, pp. 53–70.

Rice, Tara, and Kristin Stanton, 2003. "Estimating the volume of banks' payments-driven revenues," Federal Reserve Bank of Chicago, Emerging Payments, occasional papers series, No. 2003-1C.

Stiroh, Kevin J., 2004, "Diversification in banking: Is noninterest income the answer?," *Journal of Money, Credit, and Banking*, forthcoming.

_____, 2004. "Do community banks benefit from diversification?," *Journal of Financial Services Research*, Vol. 25, No. 2/3, pp. 135–160.

Wade, Will, 2003, "Wachovia unit's goal: Centralize payment biz," *American Banker*, March 4.

How do banks make money? A variety of business strategies

Robert DeYoung and Tara Rice

Introduction and summary

Banks make money many different ways. Some banks employ traditional banking strategies, attracting household deposits in exchange for interest payments and transaction services and earning a profit by lending those funds to business customers at higher interest rates. Other banks employ nontraditional strategies, such as credit card banks or mortgage banks that offer few depositor services, sell off most of their loans soon after making them, and earn profits from the fees they charge for originating, securitizing, and servicing these loans. In between these two extremes lies a continuum of traditional and nontraditional approaches to banking—focusing on local markets or serving customers nationwide; catering to household customers or business clients; using a brick-and-mortar delivery system or an internet delivery system; and so on.

This panoply of business strategies is a relatively new development in the U.S. banking industry, made possible by deregulation, advances in information technology, and new financial processes. To date, academic economists have performed very little systematic analysis of the relative profitability, riskiness, or long-run viability of these different banking business models. Academic studies of bank performance tend to focus on issues of regulatory concern (for example, capital adequacy, bank insolvency) or investor concern (for example, the reaction of bank stock prices to bank mergers) rather than broader questions of competitive strategy. Moreover, many so-called studies of banking business strategies focus myopically on banking company size. Although banks of different sizes often do different things in different ways, size is a poor proxy for strategy: It assumes that the banking strategy space has only one dimension; it assumes that a bank's size always constrains its choice of a business model; and it assumes that two banks of the same size always use the same strategy. As we demonstrate in this article, none of these assumptions are accurate. Moreover, failing to recognize this can result in a misleading analysis of bank performance.

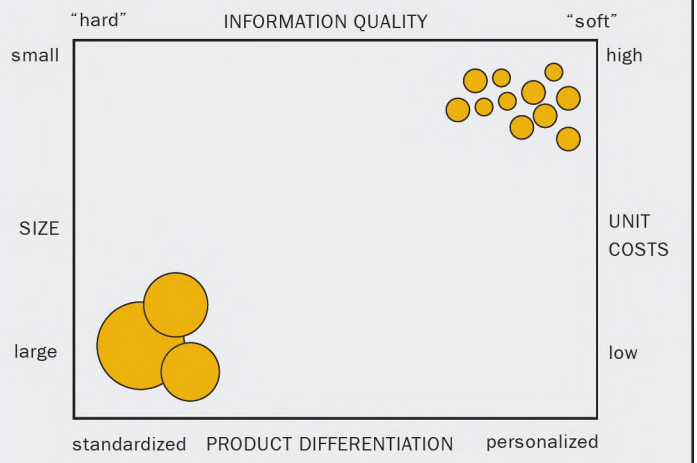
This is the second of two companion pieces on "How do banks make money?" appearing in this issue of *Economic Perspectives*. In the first article, we focus on the remarkable increase in noninterest income at U.S. commercial banks during the past two decades, the regulatory and technological catalysts for this historic change, and how this newfound reliance on noninterest income can affect bank performance. In this article, we explain how deregulation and technological change have encouraged U.S. commercial banks to become less like each other in virtually all aspects of their operations—including the generation of noninterest income—and how the resulting divergence in banking strategies has affected the financial performance of these companies. We define a variety of banking business strategies based on differences in product mix, funding sources, geographic focus, production techniques, and other dimensions, and examine the financial performance of established U.S. banking companies that used these strategies from 1993 through 2003. While we recognize that bank size can have implications for strategic choice and financial performance, we do not use bank size to define any of the strategy groups.

We draw a number of conclusions about "how banks make money" and how this may matter for the future of the banking industry. First, we find substantial differences in profitability and risk across the various banking strategy groups. Importantly, low profitability does not necessarily doom a banking

Robert DeYoung is a senior economist and economic advisor and Tara Rice is an economist in the Economic Research Department of the Federal Reserve Bank of Chicago. *The authors thank Carrie Jankowski and Ian Dew-Becker for excellent research assistance and Rich Rosen for helpful comments.*

FIGURE 1

A strategic map of the banking industry



Source: DeYoung, Hunter, and Udell (2004).

strategy. High average return strategies like corporate banking tend to generate high amounts of risk, while low average return strategies like community banking tend to generate less risk; thus, on a risk-adjusted basis, both high-return and low-return strategies may be financially viable. Second, we find that very small banks operate at a financial disadvantage regardless of their competitive strategy. This suggests that the number of very small U.S. banking companies is likely to continue to decline in the future. However, our analysis suggests that the business strategies typically associated with small banks are financially viable when practiced by "larger-than-average small banks," and we stress that under some circumstances even very small banking companies can succeed. Third, we find some evidence that banking companies without discernable competitive strategies tend to perform poorly, as do banks that employ traditional banking strategies without embracing efficient new production methods. Both of these findings are consistent with fundamental precepts of good strategic management.

Banks have become less alike

Prior to the 1990s, banking companies in the U.S. were relatively (though not completely) homogeneous. In contrast, today's commercial banking companies are substantially different from each other in terms of size, geographic scope, organizational structure, product mix, funding sources, service quality, and customer focus. This strategic diversity is a byproduct of two decades of deregulation and technological change—dramatically disruptive changes in the structural underpinnings of our financial system, which we address in detail in the two sections that follow.

DeYoung, Hunter, and Udell (2004) argue that two generic banking strategies have emerged from the fog of deregulation and technological change. This is illustrated in figure 1, which describes the strategic aftermath of deregulation and technological change using four parameters: bank size, bank unit costs, product differentiation, and information quality. The vertical dimension in the map measures bank size, with large banks at the bottom and small banks at the top. Large size allows banks to achieve low unit costs through scale economies. The horizontal dimension measures the degree to which banks differentiate their products and services from those of their competitors. To provide personalized financial services, banks must have non-quantifiable, or "soft," information about their customers. In this framework, banks select their business strategies by combining a high or low level of unit costs with a high or low degree of product differentiation. The positions of the circles indicate the business strategies selected by banks and the relative sizes of the circles indicate the relative sizes of the banks.

The first of these two generic strategies, represented by the small bubbles in the upper right-hand corner of the map, is a traditional banking strategy. Small banks operating in local markets develop close relationships with their customers, provide value to depositors through person-to-person contact at branch offices, and make "relationship loans" to informationally opaque borrowers (for example, small businesses) that do not have direct access to financial markets. Although these locally focused banks operate with relatively high unit costs, they can potentially earn high interest margins: They pay low interest rates to a loyal base of core depositors and they charge high interest rates to borrowers over which they have market power due to information-based switching costs. These banks earn fee income mainly through service charges on their deposit accounts.

The second of these two generic strategies, represented by the large bubbles in the lower left-hand corner of the map, is a nontraditional banking strategy. Large banks take advantage of economies of scale in the production, marketing, securitization, and servicing of "transaction loans" like credit cards and home mortgages. These banks operate with low unit costs,

but they tend to earn low interest margins because the loans they produce are essentially financial commodities that are sold in highly competitive markets. Large amounts of noninterest income (for example, fees from loan origination, securitization, and servicing) are essential for this model to be profitable. Note that this approach to commercial banking became possible only after geographic deregulation allowed banks to achieve larger scale and after new technologies (for example, credit scoring models, asset securitization) permitted banks and other financial institutions to create transaction loans.

It is important to observe that the highly stylized banking strategies portrayed in figure 1 are characterized not just by differences in bank size, but more fundamentally by differences in customer preferences, information quality, pricing structures, and production techniques. As such, this analysis implies that there is a rich diversity of potentially profitable business strategies for serving retail and commercial banking customers. More fundamentally, it implies that the banking companies pursuing those strategies should have grown less like each other than in the past.

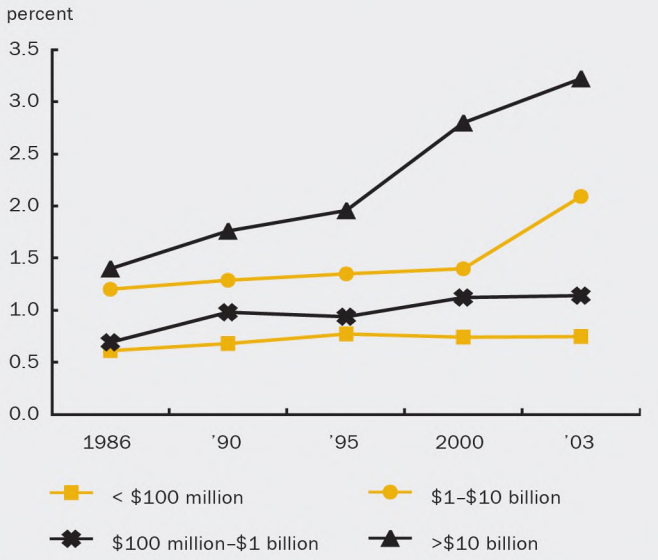
Indeed, there is evidence that they have. Figures 2 and 3 illustrate two of the dimensions across which U.S. banking companies have become less alike since 1986. (The data used to construct these figures are described in the previous article. See table 2 of that article and the associated text.)

Figure 2 shows that the intensity of noninterest income at banking companies of different sizes—very small (with inflation-adjusted assets less than $100 million), small ($100 million to $1 billion), mid-sized ($1 billion to $10 billion), and large (greater than $10 billion)—has systematically diverged over the past two decades. Noninterest income has become more important on average for banks of all four sizes; however, it has increased by only about 25 percent for the smallest banking companies while more than doubling for the largest banking companies. These trends are consistent with the emergence of the strategic dichotomy depicted in figure 1.

Banks have also grown less alike in the way they fund their loans and other investments. Figure 3 displays the distribution of transaction deposits to assets for banking companies in 1986 and 2003.[1] This distribution has flattened out over time, but not symmetrically. On the one hand, there has been a considerable displacement to the left, indicating that transaction deposits have become a less important funding source for many banking companies. On the other hand, the stable right-hand side of the distribution indicates that transaction deposits have remained a core source of funding for many other commercial banks. Again, this is consistent with the strategic dichotomy illustrated in figure 1.

FIGURE 2

Noninterest income to assets



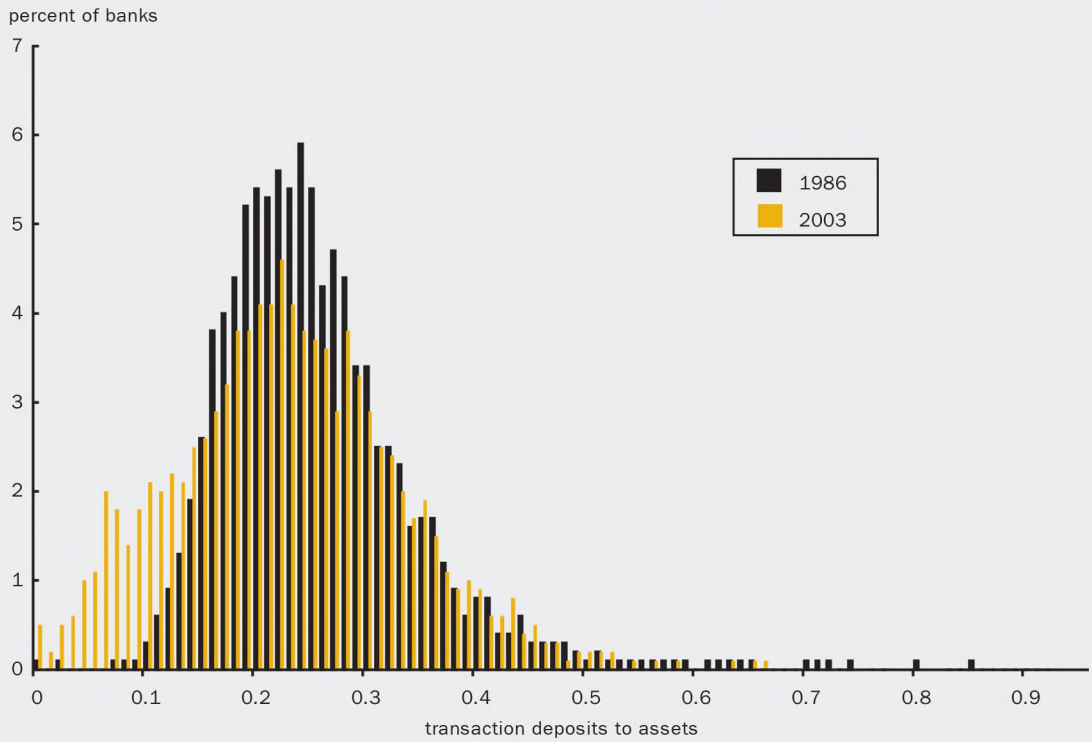
Although some of the growing dissimilarities across banking companies are clearly associated with growing differences in bank size, there are rich strategic differences across commercial banking companies that have little to do with size. As we show later in this article, these strategic differences lead to substantial heterogeneity in the financial performance of banking companies. But before we get to that analysis, we need to review the fundamental changes to the banking environment that allowed banking companies to grow so dissimilar in the first place.

Deregulation and banking business strategies

Over the past 25 years, U.S. commercial banking has been transformed from a heavily regulated industry, in which banks were prohibited from competing with each other, to a largely deregulated industry, in which commercial banks compete vigorously among themselves, as well as with investment banks, securities firms, and insurance companies. This historic industry deregulation, in conjunction with dramatic

FIGURE 3

Divergence in transactions deposits to assets, histogram for 1986 and 2003



advances in banking technology, laid the groundwork for new business strategies at commercial banks.

Deregulation has transformed almost every facet of the banking industry. It has been pro-competitive by allowing banks to expand into neighboring cities and states, to offer financial products and services that had previously been reserved for non-bank financial institutions, and to set deposit interest rates according to market forces. Deregulation has been pro-efficiency: It encouraged scale economies by allowing banks to grow larger; cost and revenue synergies by allowing banks to broaden their product lines; and operational efficiencies by exposing banks to increased market competition. And deregulation has been pro-technology by allowing banking companies to attain the large size necessary to fully benefit from declining cost technologies such as credit scoring and asset securitization, to launch mass-market advertising, and to better reduce risk via diversification.

Kane (1996), Kroszner and Strahan (1997, 1999), and others argue that it was the behavior of banking companies themselves that brought deregulation. Banks routinely circumvented regulatory constraints on geographic and product market expansion in the years prior to deregulation, and these commentators argue that deregulation was the optimal government response because the relative cost of maintaining the restrictions to one interest group (for example, large banking companies) had became less than the relative benefit of maintaining the restrictions to other interest groups (for example, small local banks that had been protected from competition).

Deregulation has been a continuous and ongoing process since the mid-1970s. Spong (2000) and DeYoung, Hunter, and Udell (2004) offer in-depth treatments of the evolution of banking and financial regulations over the past quarter-century and the impact of those changes on the structure, strategies, and performance of commercial banks. We limit our discussion here to just three deregulatory acts that have proven to be especially influential for the competitive strategies of commercial banking companies.

The *Depository Institutions Deregulation and Monetary Control Act* of 1980 sought to equalize the competitive positions of commercial banks and thrift institutions. Among other things, the act expanded

the lending powers of thrift institutions to better match those of commercial banks; increased deposit insurance coverage to $100,000 for all insured depository institutions; authorized new products such as NOW (negotiable order of withdrawal) accounts nationwide; and required the Federal Reserve to price its financial services (for example, check clearing) and make those services, as well as the discount window, available for all commercial banks and thrifts. But for commercial banking strategies, the most fundamental and far-reaching consequence of this act was the six-year phase out of Regulation Q.

Since the 1930s, Regulation Q had limited the interest rates that banks could pay their customers on time and savings deposits. Whenever competition for deposits increased—for example, if a new deposit-taking institution entered the local market or if alternative investment vehicles became more attractive than bank deposits—banks could not respond by paying higher rates to their depositors. Instead, banks compensated depositors for below-market interest rates by giving them a "bundle" of related services (for example, check printing, safety deposit boxes, travelers' checks) free of charge. This situation was extremely inefficient—banks could, at best, only respond crudely to changes in deposit market conditions and, in a world of bundled pricing, banks had little incentive to develop innovative deposit services for which they could charge customers.

Since the phase-out of Regulation Q, banks have gradually reduced bundled pricing in favor of charging explicit fees for individual retail deposit products and adjusting deposit interest rates up and down to reflect market conditions. Free to charge explicit fees for depositor services, banks had greater incentives to offer new deposit-related products such as money-market mutual funds, online bill pay, and overdraft protection. Free to pay market rates for deposits, efficiently run banks that could use deposits the most productively became able to bid those funds away from less efficient banks.

The *Riegle–Neal Act* of 1994 eliminated nearly all barriers to the geographic expansion of banking companies across state boundaries. This federal measure put the finishing touch on over 20 years of piecemeal deregulation by the states, which began in the mid-1970s with the removal of existing restrictions on in-state branching in a handful of individual states and culminated with a number of multi-state compacts that allowed banking companies to own and operate affiliates in other states. By sweeping away most federal restrictions and remaining state restrictions on interstate banking and branching, the Riegle–Neal Act gave banking companies the freedom to enter new states either by purchasing existing banking franchises or by opening new branches and allowed multibank holding companies to consolidate their separate banking affiliates into systems of branch offices.

These changes had their most visible impact on the structure of the banking system. A wave of interstate mergers and acquisitions has created a handful of nearly nationwide banking companies (for example, Bank of America, Citibank, J. P. Morgan–Chase), as well as a second tier of superregional banking companies (for example, Wells Fargo, Fifth Third, Wachovia), most of which exceed the size of the largest pre-Riegle–Neal banking companies. This geographic expansion has, in turn, provided new opportunities for both large and small banking companies to improve their operational efficiency. Duplicative back office systems (such as payroll and accounting) and organizational expenditures (separate boards of directors, bank examinations, and so on) could be eliminated by consolidating individual banks into networks of branches. Automated, information-intensive applications like credit scoring and asset securitization became more cost effective as business volume increased. Entry by large, out-of-state banking companies has increased competitive rivalry in local banking markets and created incentives for increased efficiency at local banks (DeYoung, Hasan, and Kirchhoff, 1998).

But the economies made possible by increased bank size can come at a cost, especially for large retail banks. For example, automated credit card lending and online bill-paying are low-cost ways to produce large volumes of traditional banking services, but these processes have changed the nature of retail banking from a high-touch, relationship-based service to an arms-length, financial commodity business. DeYoung, Hunter, and Udell (2004) argue that this change has had a profound influence on the business strategies of large banking companies: Because commodities do not command high margins, large banking companies may come to rely on marketing and the creation of brand images to support prices (much like other large consumer product companies). And although geographic deregulation has put community banks at a cost disadvantage relative to large banking companies, the small size of community banks can work to their strategic advantage by allowing them to provide the personal service for which deposit customers are willing to pay higher prices (or accept lower interest rates) and for which small business customers are willing to pay higher interest rates.

The *Gramm–Leach–Bliley Act* of 1999 expanded the permissible activities of commercial banking

companies. Formally, Gramm–Leach–Bliley (GLB) repealed sections 20 and 32 of the Glass–Steagall Act of 1933, a Depression-era law that effectively prohibited commercial banks from engaging in investment banking activities. In practice, GLB allows well-run commercial bank holding companies to engage in securities underwriting, securities brokerage, mutual fund services, financial advisement, and related activities without limitation, so long as these activities are conducted in a separate affiliate of the holding company. For well-run banks with federal charters, GLB permits separately capitalized financial subsidiaries.

Similar to the Riegle–Neal Act, GLB was preceded by a series of regulatory rulings during the 1990s that incrementally relaxed restrictions on banking powers. For example, the Office of the Comptroller of the Currency granted national banks to power to sell insurance from offices in small towns, and the Federal Reserve partially relaxed the limitations on the amount of revenue a banking company could generate in its Section 20 securities subsidiaries. But the new product powers granted by GLB made a bigger difference by completely relaxing the restrictions on the permissible volumes of nonbanking activities and by allowing commercial banks to engage in completely new activities such as merchant banking.

Some commercial banks now provide "one-stop-shopping" for the typical retail customer, including mortgage loans, credit cards, checking accounts, investment products and advice, and insurance products. Similarly, some commercial banks now offer a full range of financing options to their corporate customers, including loans, debt underwriting, and stock underwriting. In either case, GLB allows commercial banks to expand their traditional banking business into less traditional financial service areas by leveraging their existing distribution networks as well as the proprietary information they have gleaned over the years about their retail and corporate customers.

Technological innovation and banking business strategies

Financial services is among the industries that have been most transformed by technological change. Advances in information flows, communications infrastructure, and financial markets have dramatically altered the way in which banks assess the creditworthiness of their loan customers, service their deposit customers, process payments, and produce and distribute nearly all of their other products and services. Coupled with the effects of industry deregulation, technological advances have led to substantially increased competition in the financial marketplace as both banks and their nonbank rivals have become continuous innovators, forever attempting to improve and expand the number and variety of financial products and services that they offer.

To be sure, technological change would have occurred in the banking industry even in the absence of deregulation. But deregulation sped the application of new technologies by allowing banks to achieve the scale necessary to use new technologies efficiently and, by enhancing competition, deregulation provided banks with incentives to adopt and adapt these new technologies. As discussed above, this process also worked in the opposite direction, with technological advance speeding the progress of deregulation. As new technologies increased the efficiency of large-scale banking and created synergies between traditional and nontraditional banking products, the industry and its advocates were able to bring pressure to break down the barriers to geographic expansion. This included bold circumvention of existing legal constraints on geographic and product market expansion, the most famous of which was the 1998 merger of banking giant Citibank with insurance giant Travelers, more than a year before the passage of the Gramm–Leach–Bliley Act in 1999.

Technological changes in the banking industry can be roughly separated into two categories: improvements in data processing and communications technologies and the emergence of entirely new financial instruments, markets, and production processes. The former has allowed financial information to flow more quickly, accurately, and cheaply; the latter largely reflects the manner in which banking companies and their competitors have exploited these new information flows. Together, these phenomena have played key roles in the evolution of bank business strategies and the ways that banks make money. We offer three examples here.

Payment services

Faster information flows have transformed the manner in which banks provide payment services to their customers. The development and expansion of electronic payment channels and instruments have permitted banks to offer their deposit customers unprecedented levels of convenience, often at lower costs. For example, today about 34 percent of household payments are made using electronic channels like debit cards, credit cards, and automated bill pay; as recently as 1990 only about 15 percent of household transactions were electronic, with the remaining 85 percent made with cash and checks (HSN Consultants, Inc., 2002).

The reduction in use of the physical paycheck is testimony to the important role of transactions made

through the Automated Clearing House (ACH). ACH not only makes direct deposit of household wages possible, but it facilitates automated online bill pay for households and businesses, in addition to other recurring transactions. Retail business customers benefit from electronic lockbox services and check truncation, and the recently passed Check 21 legislation will accelerate these changes in our financial infrastructure by requiring banks to accept "substitute checks," which can be transmitted as electronic images. And for those who wish to make old-fashioned cash transactions, financial information that flows through ATM (automated teller machine) networks has made access to cash more convenient, while generating fee income for banks and creating an entirely new financial service sector for nonbank owners of ATMs.

Online brokerage

A more specialized application of financial information technology is online discount brokerage. Online brokerage of any sort was obviously not possible prior to the invention of the internet, and the discount brokerage model fits well with this distribution channel. This application reduces production costs two different ways: potential scale economies from operating on a nationwide basis and potential reductions in overhead expenses by targeting "do-it-yourself" customers. (For these customers, less personal service ironically translates into greater convenience.) Because this product is offered in a very competitive marketplace, online discount brokerage firms like Charles Schwab and E*Trade must pass a large portion of these savings on to their customers in the form of lower transaction fees.

Along with other changes in the retail financial landscape—like the widespread adoption of mutual fund investing and the shift to defined contribution pension plans—the emergence of discount brokerage firms has increased the competition for household savings and investments. In response, most large retail banks now offer some version of online brokerage to help retain retail depositors.

Intermediation

Banks have traditionally earned most of their profits by intermediating between parties that have excess liquidity (depositors) and parties that need additional liquidity (borrowers). For a variety of reasons, banks historically have been better than other institutions at mitigating the informational asymmetries and other logistical problems that prevent direct finance between these parties.[2] But advances in information processing and financial markets have greatly reduced banks' comparative advantages, and the resulting "disintermediation" has changed, in some cases dramatically, the roles that banks play in credit markets.

On the consumer lending side, the advent of credit scoring models that use "hard" (that is, quantifiable) information to evaluate creditworthiness, together with the development of secondary markets for securitized loans, has changed the way that banks and other financial institutions provide credit to households (Stein, 2002). Instead of earning interest margins from holding mortgage, auto, or credit card loans in their loan portfolios, banks can earn separate fees for originating the loans, securitizing the loans, and servicing the loans, while the interest income flows to the investors that purchase the securities backed by these loans. New financial institutions—such as brokers that originate and immediately securitize home mortgages and monoline credit card and finance companies that take advantage of huge scale economies in the production, distribution, and servicing of consumer credit—have emerged to service much of the market share in consumer credit that traditionally belonged to depository institutions like banks.

On the business lending side, the introduction of high-yield ("junk") bonds, increased access to commercial paper, and other financial market developments have allowed large commercial borrowers to bypass banks in favor of direct finance. While commercial banks have lost considerable market share in commercial lending, one way that they continue to play a role in commercial finance is by charging a fee in exchange for providing the back-up lines of credit that firms need to float commercial paper. In this new technological environment, loans to small and moderate-sized businesses based on private, information-rich relationships between business people and their commercial bankers stand out as one of the last types of loans that are still produced in the traditional intermediation fashion.

Business strategies at banking companies

A simple and often-employed method for comparing the performance of different banking strategies is to separate banking companies by size. As we have seen, scale is clearly important: The scale of a large banking company gives it access to low-unit-cost marketing and production techniques, while the scale of a small banking company allows it to build person-to-person relationships with its customers. But economies of scale is not the only dimension across which banking companies vary strategically. Moreover, we assume that achieving a large scale, a medium scale, or a small scale is not the main objective of a banking

company; rather, it is to earn a rate of return commensurate with the risk to which owners of the bank are exposed. In pursuit of high risk-adjusted earnings, banking companies choose from among many banking strategies, some of which can be practiced by small banks as well as large banks.

For the purposes of this study, we define eight distinct banking business strategies based on differences in product mix, location, production techniques, and other characteristics across U.S. banking companies: traditional banking, nontraditional banking, private banking, agricultural banking, corporate banking, local community focus, payment transactions, and a diversified banking strategy. The procedures we use to define these strategy groups, and to assign banking companies to these groups, are presented below and are not highly scientific. We used our informed judgment to select a short list of characteristics that one would expect to find at banks that used each of these business strategies and we set arbitrary numerical thresholds for each of those characteristics above or below which banking companies would be included in, or excluded from, each strategy group. It is important to note that we did not use bank size to define any of these eight strategy groups and, as a result, each strategy group includes banking companies of different sizes. (For comparative purposes, we also define a number of groups based purely on bank size and bank growth rates.)

Banking companies were eligible for assignment to one or more of these strategy groups if they were at least ten years old in 1993,[3] were still operating in 2003, were domestically owned, and had positive amounts of loans, transaction deposits, deposits insured by the Federal Deposit Insurance Corporation (FDIC), and equity capital in both 1993 and 2003. A total of 1,281 banking companies met these eligibility conditions. We selected the 1993–2003 period because it began after the passage of the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991 and because it was long enough to adequately observe the variability of banking company returns over an entire business cycle. We drew the data for our analysis chiefly from the *Reports of Condition and Income* (call reports), Federal Reserve Board FR Y-9C reports, the Federal Reserve Board National Information Center's (NIC) structure database, the Federal Deposit Insurance Corporation's *Summary of Deposits* database, and the Center for Research on Stock Prices (CRSP) database. We express all data in thousands of year 2003 dollars, unless otherwise indicated.

The eight business strategies are not meant to be fully exhaustive of all the competitive strategies being used by banking companies today. Moreover, we defined the strategy groups tightly: Over half (758 out of 1,281) of the eligible banking companies were not assigned to any strategy group. Although we did not design the strategy groups to be mutually exclusive, only about 10 percent (123) of the 1,281 banking companies fell into more than a single group; of these, just 28 banking companies were assigned to three or more strategy groups, and just two banking companies were assigned to four strategy groups.

The *traditional banking* group contains 117 banking companies; 2003 assets averaged about $242 million and ranged from $10 million to $1.7 billion. To be included in this strategy group, banking companies had to be portfolio lenders that did not securitize any assets in either 1993 or 2003, and their ratios of core deposits to assets, loans to assets, and net interest income to operating income all had to rank higher than the 25th percentile among our sample of 1,281 banking companies in both 1993 and 2003.

The *nontraditional banking* group contains 29 banking companies; 2003 assets averaged about $140 billion and ranged from $590 million to $771 billion. To be included in this strategy group, banking companies had to securitize at least some assets in both 1993 and 2003; rank lower than the 25th percentile in our sample in terms of both deposits to assets and net interest income to operating income; and rank above the 75th percentile in terms of the asset value of letters of credit issued to assets. This group includes many of the nationally recognized commercial banking companies (for example, Bank of America, J. P. Morgan–Chase, Wachovia, Wells Fargo), as well as a number of superregional (for example, Fifth Third, National City, Suntrust) and regional commercial banks (for example, First Tennessee, Marshall & Ilsley, Regions Financial).

The *private banking* group contains 11 banking companies; 2003 assets averaged about $25 billion and ranged from $550 million to $92 billion. To be included in this strategy group, banking companies had to rank above the 99th percentile in terms of fiduciary income to operating income in 1993 and 2003. Some of the better known companies in this group are Northern Trust, State Street, Bank of New York, and Mellon Financial.

The *agricultural banking* group contains 96 banking companies; 2003 assets averaged $108 million and ranged from $4 million to $1.2 billion. To be included in this strategy group, banking companies had to rank

above the top 90th percentile in terms of agricultural production loans to total loans in both 1993 and 2003.

The *corporate banking* group contains 14 banking companies; 2003 assets averaged about $74 billion and ranged from $729 million to $327 billion. To be included in this strategy group, investment banking activities had to generate at least 1 percent of a bank's operating income in 2003; the bank had to rank above the 75th percentile in commercial loans to total loans in both 1993 and 2003; and the bank had to rank above the 50th percentile of the sample in terms of demand deposits to total deposits and the asset value of letters of credit issued to assets in both 1993 and 2003. Some of the companies included in this group are Bank One (before its acquisition by J. P. Morgan–Chase), Commerce Bancshares, FleetBoston (before its acquisition by Bank of America), Huntington Bancshares, Mellon Financial, PNC Financial, and U.S. Bancorp.

The *community focus* group contains 151 banking companies; 2003 assets averaged about $268 million and ranged from $8 million to $4.1 billion. Companies in this strategy group generated at least half of their deposits from a one-county area and ranked above the 50th percentile in core deposits to assets and loans to assets, in both 1993 and 2003.

The *transaction services* group contains 96 banking companies; 2003 assets averaged about $1.6 billion and ranged from $8 million to $46 billion. Banking companies in this strategy group ranked above the top 75 percent of banking companies in terms of both payment-related income associated largely with checking transactions (service charges on deposits plus foregone interest revenue on deposits) and payment-related income not necessarily associated with checking transactions (estimated payment-related fees from ATM, fiduciary, and credit card activities) as a percentage of operating income in 2003.

The *diversified banking* group contains 97 banking companies; 2003 assets averaged about $1.6 billion and ranged from $160 million to $26 billion. This strategy group includes banking companies that do not specialize in any of the areas described above but participate to at least some extent in each of those areas. To be included in this strategy group, banks had to rank between the 10th and 90th percentiles among the 1,281 eligible banks in terms of service charges to assets, other (non-service charge) noninterest income to assets, net interest income to assets, home mortgage loans to total loans, commercial real estate loans to total loans, and consumer loans to total loans in both 1993 and 2003. Moreover, these banks had to rank above the 90th percentile in terms of commercial loans to total loans and below the 90th percentile in agricultural production loans to total loans, in both years.

In addition to these eight largely activities-based strategy groups, we defined five purely *size-based strategy groups*: assets less than $100 million (541 banks); assets between $100 million and $500 million (303 banks); assets between $500 million and $1 billion (59 banks); assets between $1 billion and $10 billion (89 banks); and assets greater than $10 billion (29 banks). We applied these size thresholds to the assets of each banking company twice: In 2003 we applied them to actual 2003 asset values, and in 1993 we applied them to 1993 asset values that had been adjusted upward to account for industry asset growth and inflation between 1993 and 2003. We also defined two strategy groups based on the asset growth rates. The geographic deregulation of U.S. banking markets in the late 1980s and 1990s created unparalleled opportunities for U.S. banking companies to grow, either by making acquisitions or by growing internally. The *mergers* (external growth) group contains 17 banking companies, with 2003 assets averaging $143 billion in a range from $514 million to $771 billion. These banking companies grew at an inflation-adjusted rate of 250 percent or more between 1993 and 2003, and at least 25 percent of this increased size was attributable to assets acquired in mergers. The *growers* (internal growth) group contains 85 banking companies, with 2003 assets averaging $2.8 billion and ranging from $47 million to $88 billion. These banking companies grew at an inflation-adjusted rate of 250 percent or more between 1993 and 2003 without making any major acquisitions.

Finally, we defined a *no-strategy* group. This group contains 113 banking companies that did not qualify for any of the eight main strategy groups in both 1993 and 2003. (Note that the no-strategy group does not include banking companies that "switched" strategies, that is, banks that qualified for one of the eight main strategy groups in 1993 and qualified for a different strategy group in 2003. The financial performance of these banks would likely have been impacted by the costs of transitioning from one business strategy to another.)

Financial performance of different business strategies

We used quarterly accounting data to calculate three financial performance measures for each of the 1,281 banking companies in our 1993–2003 dataset: The profitability of each bank is the annualized average return on equity (ROE) over the 44 quarters from 1993 through 2003. The riskiness of each bank is the

annualized standard deviation of quarterly ROE over that period. The risk-adjusted return of each bank, also known as the Sharpe ratio, is the annualized quarterly ROE minus the annualized interest rate on 90-day Treasury bills, divided by the annualized standard deviation of quarterly ROE.[4] In addition, for the 157 banking companies in our dataset that were publicly traded, we used weekly stock prices to calculate market-based analogs of these three financial performance measures.

Table 1 displays summary statistics (means and standard deviations) for all of our performance measures. We note that our performance measures are observed ex post—that is, they reflect actual rather than expected revenues and expenses—and as such they are just proxies for investors' expectations of future returns, upon which finance theory is based. We also note that our dataset excludes banking companies that were acquired or failed between 1993 and 2003, and as a result the performance measures for the "surviving" companies that populate our dataset may be biased. For example, banks that practice especially risky strategies will be more likely to fail, all else being equal, so the average ROE for a high-risk strategy group may be biased upward.

The quarterly accounting-based returns exhibit considerably less variation over time—and as a result, substantially higher risk-adjusted profits—than the weekly stock market returns. This difference is likely due to three factors: accounting conventions that affect the valuation of assets and the way that expenditures are recognized over time; changes in relevant information and investor expectations that are priced by the stock market but not included in backward-looking accounting statements; and the different frequencies over which we observe the accounting data and the market data (quarters versus weeks).[5] Also note that the accounting-based returns are substantially lower on average than the stock market-based returns. The most likely explanation is that publicly traded companies with low returns are likely to become takeover targets and drop out of our sample, while closely held private companies (often small banks) with low returns are more likely to continue to operate independently.

TABLE 1

Accounting-based and market-based financial performance measures

	Quarterly accounting ROE 1,281 companies	Weekly stock market returns 157 companies
	mean (standard deviation)	
Return	0.1199 (0.0785)	0.1726 (0.0708)
Risk	0.0337 (0.0413)	0.2813 (0.0721)
Risk-adjusted return	4.0646 (3.5873)	0.4636 (0.2094)

Notes: ROE is return on equity. Performance measures are observed ex post. Banking companies that were acquired or failed are not included. Measures are not comparable across columns.

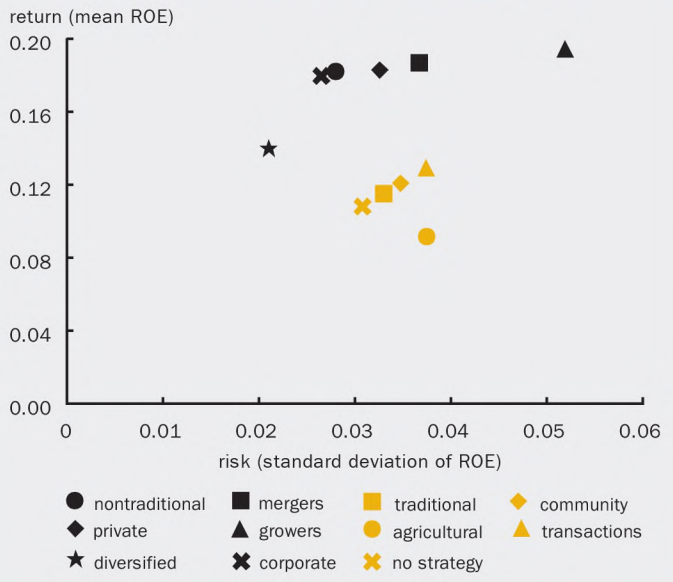
Accounting-based financial performance

Figure 4 plots the average accounting-based return and risk measures for each of the strategy groups. These average risk-return profiles fall into two clusters. One cluster of strategies (diversified, corporate, nontraditional, private banking, mergers, and growers) forms an arc of risk–return combinations consistent with the fundamental principle of finance that markets reward risk-taking with higher returns, but that the returns to risk-taking are diminishing. (This arc is not a representation of the "efficient" risk-return frontier, because we have plotted it based on the average risks and average returns of the banks in each strategy group.)[6] Moving from left to right on the graph—from low-risk–low-return strategies to higher-risk–higher-return strategies—these strategy groups line up in an economically sensible order. Not surprisingly, the diversified strategy has the lowest (ex post) risk position. The corporate, nontraditional, and private banking groups come next, with increasingly higher levels of risk (and associated higher returns) that are roughly consistent with the increasing reliance of the banks in these groups on noninterest income (DeYoung and Roland, 2001).

The highest risks and the highest returns, on average, are generated by banking companies that grew quickly during the sample period by either external means (the mergers group) or internal means (the growers group). For the merging banks, accounting earnings are likely to be volatile because of accounting charges taken during the post-merger transition period. For the growing banks, this volatility is likely related to several different phenomena: the temporary excess capacity in physical plant necessary to grow a bank by opening new branch locations; a slippage in credit quality that often occurs when banks attempt to grow their loan portfolios quickly; and the purchase of expensive time deposit funding to which these banks often must resort to finance fast asset growth. The high accounting earnings also have a number of plausible explanations. On the one hand, profitable

FIGURE 4

Average book returns and risk for strategy groups



banks are better able to generate the large amounts of internal funds to make the repeated purchases or investments necessary to expand rapidly. On the other hand, the data simply may indicate that merger-based and growth-based strategies tended to pay off during the 1990s (Calomiris and Karceski, 1998). Finally, the returns for the high-risk growers strategy may be biased upward to the extent that unsuccessful fast-growing banks that failed are not in our dataset.

A second cluster of strategies (traditional, community focus, transactions, agricultural, and no strategy) lies well below the risk-return arc. The returns generated by the banks using these strategies do not appear to be high enough to compensate bank owners for the risks they are taking—in other words, the data suggest that these are not economically viable banking strategies, and these strategies and the banking companies that use them could disappear from the banking industry sometime in the future. But before writing off these banking strategies, we note that there is a substantial size disparity between the two clusters of banking companies: Those on the risk–return arc tend to contain larger banks, while those in the lower cluster tend to contain small banks. Is the poor average financial performance of the banks in the second cluster of strategy groups attributable to untenable banking strategies, inefficiently small bank size, or a combination of both?

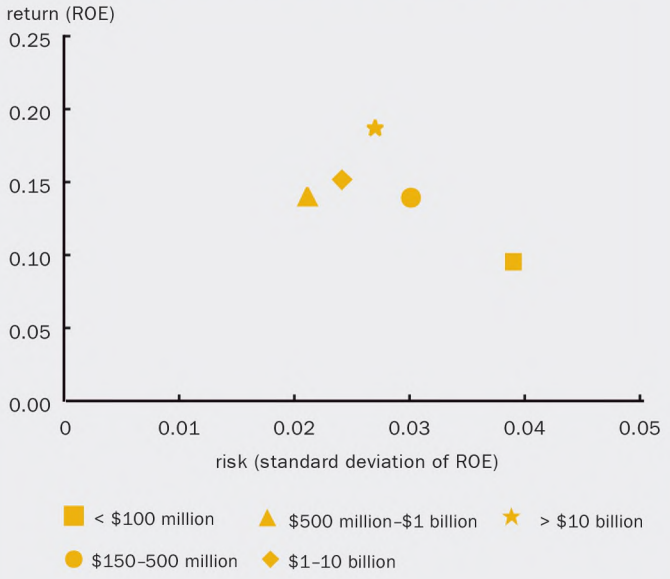
To investigate this possibility, we plotted the average accounting-based return and risk measures for each of our five purely sized-based groups. As shown in figure 5, for banking companies with assets less than $500 million increased size unambiguously improves (ex post) financial performance—that is, returns increase without having to accept more risk—while for banking companies larger than $500 million increased returns are attainable on average only by accepting increased risk. This crude analysis is consistent with several findings in the bank scale economy literature. In general, this literature finds that even relatively large banking companies can expect to reduce per-unit costs by growing larger. Berger, Demsetz, and Strahan (1999) provide a relatively recent review of this literature. However, Evanoff and Israilevich (1991) and Berger and Humphrey (1991) demonstrated that the bulk of these per-unit cost improvements are captured at relatively small bank size—that is, average costs decrease with bank size but at a rapidly diminishing rate. Other studies have found that banking companies that grow larger tend to take on increased risk (for example, Demsetz and Strahan, 1997; Hughes, Lang, Mester, and Moon, 1996), consistent with the patterns for the larger banking companies in figure 5.

Although this demonstration of the risk–return effects of increased banking company scale is admittedly crude, applying these findings to our analysis produces stark and economically sensible results. In figure 6 we re-plot the average risk profiles of the banking strategy groups after removing companies with assets less than $500 million. The result is a relatively smooth arrangement of the strategy groups along the original risk–return arc from figure 4. (The average risk–return tradeoff between these strategy groups is illustrated by a quadratic ordinary least squares trend line estimated for the 11 data points shown in the figure. Again, we note that this line is based on average financial performance, and is not an "efficient risk–return" frontier.) The community focus, agricultural, and transactions strategy groups are now located on the imaginary risk–return arc and exhibit the relatively low levels of risk that are consistent with business models that rely on close customer relationships.

Although the risk–return profiles of the traditional and no-strategy groups also improved after adjusting for scale effects, these two groups still fall somewhat short of the other strategy groups. For the no-strategy

FIGURE 5

Average book returns and risk for size groups



group, the explanation may be that firms that lack strategic direction will naturally perform poorly (Porter, 1980). For the traditional group, the explanation may be that recent advances in information flows, pricing strategies, and production methods can enhance profitability, and banking companies that do not integrate these advances into their business model will operate at a disadvantage.

Transforming risk into return

Figure 6 provides reasonably compelling evidence that changing strategies would require a banking company to accept more risk in exchange for higher returns or lower returns in exchange for lower risk, on average. However, the figure does not reveal directly whether any of these risk–return tradeoffs are superior to others. In table 2 we rank each of the strategy groups shown in figure 6 by their average risk-adjusted returns, or Sharpe ratios. The Sharpe ratio can be interpreted as a measure of how well a banking company transforms risk-taking into profitability.[7]

This average performance measure divides the strategies into three subgroups. The growers have by far the worst Sharpe ratios, equal to just 5.3 on average. Despite the possible upward performance bias for this group (discussed above), banking companies that experienced rapid internal growth tended to generate low returns relative to the riskiness of this behavior. A second subgroup includes the traditional, private, agricultural, and no-strategy banks, with Sharpe ratios ranging from 6.0 to 6.7 on average. We discussed the potential shortcomings of the traditional strategy and the no-strategy groups above. The relatively poor performance of the private banking strategy group is likely explained by the large fluctuations in the stock market during the latter half of our sample period, while the small number (five) of agricultural banks in this analysis makes the poor average performance of this group difficult to interpret. The third subgroup includes the community, corporate, mergers, diversified, transactions, and nontraditional strategy groups, with Sharpe ratios ranging from 7.3 percent to 8.1 on average. The relatively good risk–return performance of these six strategies is instructive: These strategy groups are very different in terms of product mix, customer focus, production processes, funding sources, and company size. Thus, the data in table 2 suggest that a broad range of different types of banking strategies are financially viable, once banking companies have achieved at least a modicum of scale.

We performed a complete set of pair-wise tests to see which pairs of strategy groups had statistically different average Sharpe ratios and found only a few of the pairs to be statistically different. One way to interpret this result is that all of these strategic groups can be, on average, economically viable. However, it is more likely that the small number of observations in some of the strategy groups, along

TABLE 2

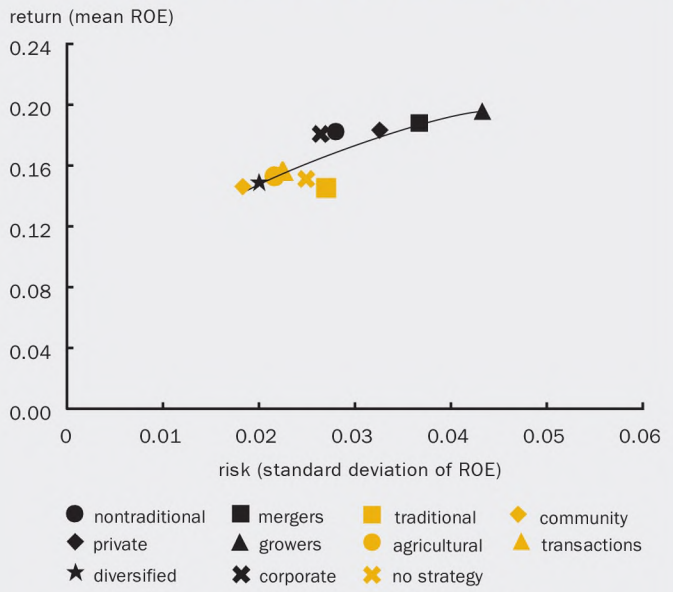
Strategy groups by average accounting-based Sharpe ratios

Rank	Strategy	Number of firms	Mean Sharpe ratio
1	Nontraditional	29	8.0621
2	Transactions	24	7.9124
3	Diversified	60	7.7869
4	Mergers	17	7.4915
5	Corporate	14	7.4682
6	Community	26	7.2875
7	No strategy	113	6.6622
8	Agricultural	5	6.4347
9	Private	11	6.3675
10	Traditional	17	6.0248
11	Growers	50	5.2830

Note: Calculations exclude banking companies with assets greater than $500 million.

FIGURE 6

Average book returns and risk for strategy groups



Note: Excludes banking companies with assets below $500 million.

with substantial noise in our estimated Sharpe ratios, is simply preventing us from finding statistical differences between most of the strategy pairs.

Market-based financial performance

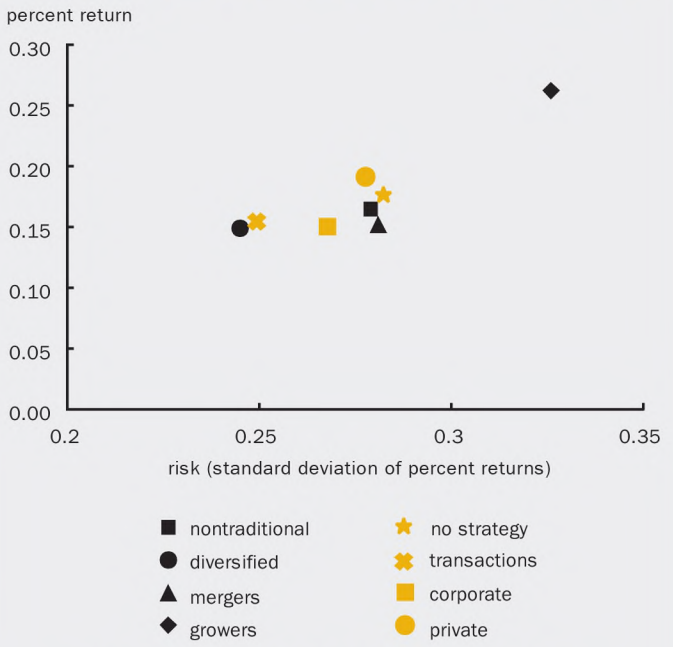
We re-plotted the risk–return averages once again, this time using stock market-based performance measures. Although we have stock market returns for only about 12 percent of the 1,281 banking companies in our sample, using these data to compare the risk–return profiles of the strategy groups provides a good robustness check on our accounting-based risk–return analysis. Stock returns reflect more information than accounting returns, and the stock prices upon which they are based are forward-looking valuations by informed investors rather than backward-looking records based on often arcane accounting rules.

Figure 7 displays the average market-based return and risk measures for the strategy groups that contained at least five publicly traded banking companies with assets greater than $500 million. The results are quite consistent with the accounting-based plots. The diversified strategy group once again defines the low-risk, low-return endpoint, and the growers group once again defines the high-risk, high-return endpoint. In between these two endpoints, six other strategies—transactions, corporate, nontraditional, private, mergers, and no-strategy—are arrayed in a risk–return ordering somewhat similar to the accounting-based ordering plotted in figure 6. Thus, we have some confidence that our accounting-based risk and return measures are providing a roughly accurate ordering of the relative risks and returns across banking strategies.

Finally, to demonstrate the large amount of variability in these data, we plotted the market-based performance measures for the individual banking companies from three strategy groups with distinctively different risk–return profiles: the diversified strategy, the nontraditional strategy, and the grower strategy. Figure 8 shows the resulting scatter plot. Although the individual data points overlap to a large extent, they do not overlap completely, and it is easy to see a rough, but positive, risk–

FIGURE 7

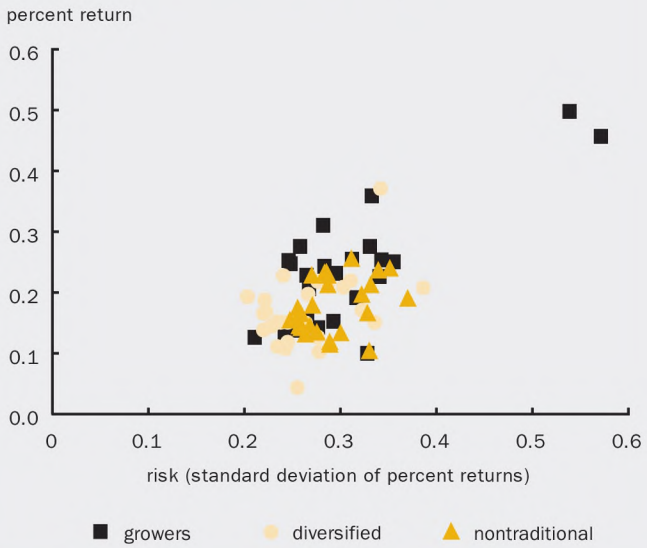
Average stock market returns and risk for strategy groups



Note: Excludes banking companies with assets below $500 million.

FIGURE 8

Stock market returns and risk for traded banking companies with assets > $500 million in three selected strategy groups



expected return tradeoff across these three strategies. The scatter plot also provides a good illustration of why it can be difficult to find statistical differences in risk-adjusted returns across strategic groups, even though the banking companies in these groups show a systematic risk–return ordering.

Conclusions and implications

We began this set of articles by asking the question "How do banks make money?" In the course of our analysis we have discussed various trends and developments in the banking industry that provide partial answers to this question. But we have also uncovered some broad themes regarding bank performance and competitive strategies that make some banks more profitable than others.

U.S. banking companies employ a wide variety of business models. For example, some are specialized and some are very broad; some have a retail focus and some have a wholesale focus; some are nationwide in scope and some are purely local; some focus on traditional commercial banking and some focus on non-bank financial services. We find substantial differences in profitability across these different strategic approaches—but we also find that high-return strategies tend to generate high amounts of risk, while relatively low-return strategies tend to generate less risk. This suggests a tradeoff between risk and return that can leave the shareholders of high-risk banks and the shareholders of low-risk banks roughly equally compensated on a risk-adjusted basis. In other words, a variety of different banking strategies—from small, locally focused community banking to large, economy-wide corporate banking—appear to be financially viable business models.

The major caveat to this conclusion is that *very small* banks tend to operate at a financial disadvantage, regardless of their business model. In order to earn a market return for their shareholders, banking companies must capture at least some of the scale economies that are available in banking production functions. Although we use an asset size threshold of $500 million to make this point in our analysis, we stress that the critical size for a banking company varies with its strategy, and even within a strategy group the critical size needed for financial viability likely varies with managerial abilities, local market conditions, and other considerations. Our analysis suggests that the number of very small U.S. banking companies is likely to continue to decline in the future. Still, there are reasons to expect that hundreds of very small banking companies will continue to exist. For example, very small banks that serve geographically isolated rural communities may remain financially viable if the lack of competition in these markets allows them to charge prices high enough to offset the cost disadvantages associated with very low scale. And, of course, very small banks whose owners are willing to operate at a relatively low rate of return in exchange for receiving personal satisfaction or providing a community service are also likely to survive in some numbers.

NOTES

[1]To check whether the shift in the distribution in figure 3 was merely due to an increase in equity capital in most banks during this period, we also examined the distribution of transaction deposits to liabilities. This distribution was nearly identical to figure 3.

[2]The function of banks as intermediaries lies at the core of a rich theoretical literature on why banks exist. See DeYoung and Rice (2004) for a short review of this literature.

[3]For bank holding companies (BHCs) and financial holding companies (FHCs), we based this threshold on the average age of the commercial banking affiliates in these multi-bank companies.

[4]The quarterly ROE data were de-seasonalized prior to these calculations, and quarters in which banking companies made large acquisitions were excluded from the calculations.

[5]To explore the extent to which the scale of the accounting-based and market-based risk and return measures differ, we recalculated the market-based measures using quarterly data. The average market-based returns fell from 0.1726 to 0.1553 and the average standard deviation of market-based returns fell from 0.2813 to 0.2579. These changes only partially closed the gap between the accounting-based and market-based measures reported in table 1. Thus, we conclude that the primary difference between the scales of the market-based and accounting-based measures lies with accounting conventions and not the frequency with which we observe the returns.

[6]We acknowledge that the average performance of banking companies that use a given strategy may not be a good comparative indicator of the potential performance of that strategy. For example, it may be the case that some strategies are attempted only by companies with very efficient management teams (in which case the average performance will be representative of the best-practice performance), while other strategies are attempted by both well-managed and poorly managed banking companies (in which case the average performance will not be representative of the best-practice performance). We plan to pursue this issue in future research.

[7]A technical point: Each of the Sharpe ratios displayed in table 2 is calculated by taking the average of the individual Sharpe ratios for the banking companies in a given strategy group. These numbers are analytically different from the Sharpe ratios implied for each of the strategy groups in figure 6, which plots the averages of the individual returns (vertical axis) and individual risks (horizontal axis) for the banking companies in each strategy group. In the figure, the Sharpe ratio is implied by the slope of a line running from about 0.043 on the vertical axis (the average risk-free rate during the sample period) through the plotted points. The two approaches are conceptually similar and result in similar rankings of the strategy groups in terms of their risk–return tradeoffs. However, the Sharpe ratio averages displayed in table 2 are superior because they directly link risk and return for each banking company, which is where the ex ante managerial decisions to trade risk for return are made.

REFERENCES

Berger, Allen N., Rebecca S. Demsetz, and Philip E. Strahan, 1999, "The consolidation of the financial services industry: Causes, consequences, and implications for the future," *Journal of Banking and Finance*, Vol. 23, February, pp. 135–194.

Berger, Allen N., and David B. Humphrey, 1991, "The dominance of inefficiencies over scale and product mix economies in banking," *Journal of Monetary Economics*, Vol. 28, pp. 117–148.

Calomiris, Charles W., and Jason Karceski, 1998, *Is the Bank Merger Wave of the 1990s Efficient? Lessons from Nine Case Studies*, American Enterprise Institute.

Demsetz, Rebecca S., and Philip E. Strahan, 1997, "Diversification, size, and risk at bank holding companies," *Journal of Money, Credit, and Banking*, Vol. 29, August, pp. 300–313.

DeYoung, Robert, Iftekhar Hasan, and Bruce Kirchhoff, 1998, "The impact of out-of-state entry on the efficiency of local banks," *Journal of Economics and Business*, Vol. 50, pp. 191–204.

DeYoung, Robert, William C. Hunter, and Gregory F. Udell, 2004, "The past, present, and probable future for community banks," *Journal of Financial Services Research*, Vol. 25, No. 2/3, pp. 85–133.

DeYoung, Robert, and Tara Rice, 2004, "Noninterest income and financial performance at U.S. commercial banks," *Financial Review*, Vol. 39, No. 1, pp. 101–127.

DeYoung, Robert, and Karin P. Roland, 2001, "Product mix and earnings volatility at commercial banks: Evidence from a degree of leverage model," *Journal of Financial Intermediation*, Vol. 10, pp. 54–84.

Evanoff, Douglas D., and Philip R. Israilevich, 1991, "Productive efficiency in banking: Econometric and linear programming evidence," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 15, July/August, pp. 11–32.

HSN Consultants, Inc., 2002, *The Nilson Report*, April, No. 761.

Hughes, Joseph P., William Lang, Loretta J. Mester, and Choon-Geol Moon, 1999, "The dollars and sense of bank consolidation," *Journal of Banking and Finance*, Vol. 23, pp. 291–324.

Kane, Edward J., 1996, "De jure interstate banking: Why only now?," *Journal of Money, Credit, and Banking*, Vol. 28, No. 2, pp. 141–161.

Kroszner, Randall S., and Philip E. Strahan, 1999, "What drives deregulation? Economics and politics of the relaxation of bank branching restrictions," *Quarterly Journal of Economics*, Vol. 114, No. 4, pp. 1437–1467.

_____, 1997, "The political economy of deregulation: Evidence from the relaxation of bank branching restrictions in the United States," Federal Reserve Bank of New York, research paper, No. 9720.

Porter, Michael E., 1980, *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, New York: Free Press.

Spong, Kenneth, 2000, *Banking Regulation: Its Purposes, Implementation, and Effects*, Federal Reserve Bank of Kansas City.

Stein, Jeremy C., 2002, "Information production and capital allocation: Decentralized vs. hierarchical firms," *Journal of Finance*, Vol. 57, pp. 1891–1921.

Index for 2004

Title & author	Issue	Pages
BANKING, CREDIT, AND FINANCE		
FDIC losses in bank failures: Has FDICIA made a difference? George G. Kaufman	Third Quarter	13–25
How do banks make money? The fallacies of fee income Robert DeYoung and Tara Rice	Fourth Quarter	34–51
How do banks make money? A variety of business strategies Robert DeYoung and Tara Rice	Fourth Quarter	52–67
ECONOMIC CONDITIONS		
The acceleration in U.S. total factor productivity after 1995: The role of information technology John G. Fernald and Shanthi Ramnath	First Quarter	52–67
Assessing the jobless recovery Daniel Aaronson, Ellen R. Rissman, and Daniel G. Sullivan	Second Quarter	2–20
Is the official unemployment rate misleading? A look at labor market statistics over the business cycle Lisa Barrow	Second Quarter	21–35
Can sectoral reallocation explain the jobless recovery? Daniel Aaronson, Ellen R. Rissman, and Daniel G. Sullivan	Second Quarter	36–49
The relationship between Hispanic residential location and homeownership Maude Toussaint-Comeau and Sherrie L. W. Rhine	Third Quarter	2–12
You can't take it with you: Asset run-down at the end of the life cycle Kate Anderson, Eric French, and Tina Lam	Third Quarter	40–54
REGIONAL ISSUES		
The state of the state and local government sector: Fiscal issues in the Seventh District Richard H. Mattoon	First Quarter	2–17
Creative destruction in local markets Jaap H. Abbring and Jeffrey R. Campbell	Second Quarter	50–60
House prices and the proposed expansion of Chicago's O'Hare Airport Daniel P. McMillen	Third Quarter	28–39
MONEY AND MONETARY POLICY		
Cyclical implications of the Basel II capital standards Anil K Kashyap and Jeremy C. Stein	First Quarter	18–31
Poor hand or poor play? The rise and fall of inflation in the U.S. François R. Velde	First Quarter	34–51
Interest rates and the timing of new production Boyan Jovanovic and Peter L. Rousseau	Fourth Quarter	2–11
In search of a robust inflation forecast Scott Brave and Jonas D. M. Fisher	Fourth Quarter	12–31

To order copies of any of these issues, or to receive a list of other publications, please telephone (312)322-5111 or write to: Federal Reserve Bank of Chicago, Public Information Center, P.O. Box 834, Chicago, IL 60690-0834. The articles are also available to download in PDF format from the Bank's website at www.chicagofed.org/economic_research_and_data/economic_perspectives.cfm