Vol. 33, No. 2

ECONOMIC REVIEW

1997 Quarter 2

FEDERAL RESERVE BANK
OF CLEVELAND

Vol. 33, No. 2

http://clevelandfed.org/research/review/
Economic Reveiw 1997 Q2

1997 Quarter 2

The Recent Ascent of Stock Prices: Can It Be Explained by Earnings Growth or Other Fundamentals?

by John B. Carlson and Kevin H. Sargent

John B. Carlson is an economist and Kevin H. Sargent is a research assistant at the Federal Reserve Bank of Cleveland. The authors would like to thank Joseph Haubrich and Peter Rupert for their helpful comments.

… how do we know when irrational exuberance has unduly escalated asset values, which then become subject to unexpected and prolonged contractions...?

We have not been able, as yet, to provide a satisfying answer to this question, but there are reasons in the current environment to keep the question on the table. Clearly, when people are exposed to long periods of relative economic tranquility, they seem inevitably prone to complacency about the future. This is understandable. We have had fifteen years of economic expansion interrupted by only one recession — and that was six years ago. As the memory of such past events fades, it naturally seems ever less sensible to keep up one's guard against an adverse event in the future. Thus, it should come as no surprise that, after such a period of balanced expansion, risk premiums for advancing funds to businesses in virtually all financial markets have declined to near-record lows.

— Alan Greenspan, February 26, 1997

A severe depression like that of 1920–21 is outside the range of probability.

—Harvard Economic Society, November 16, 1929

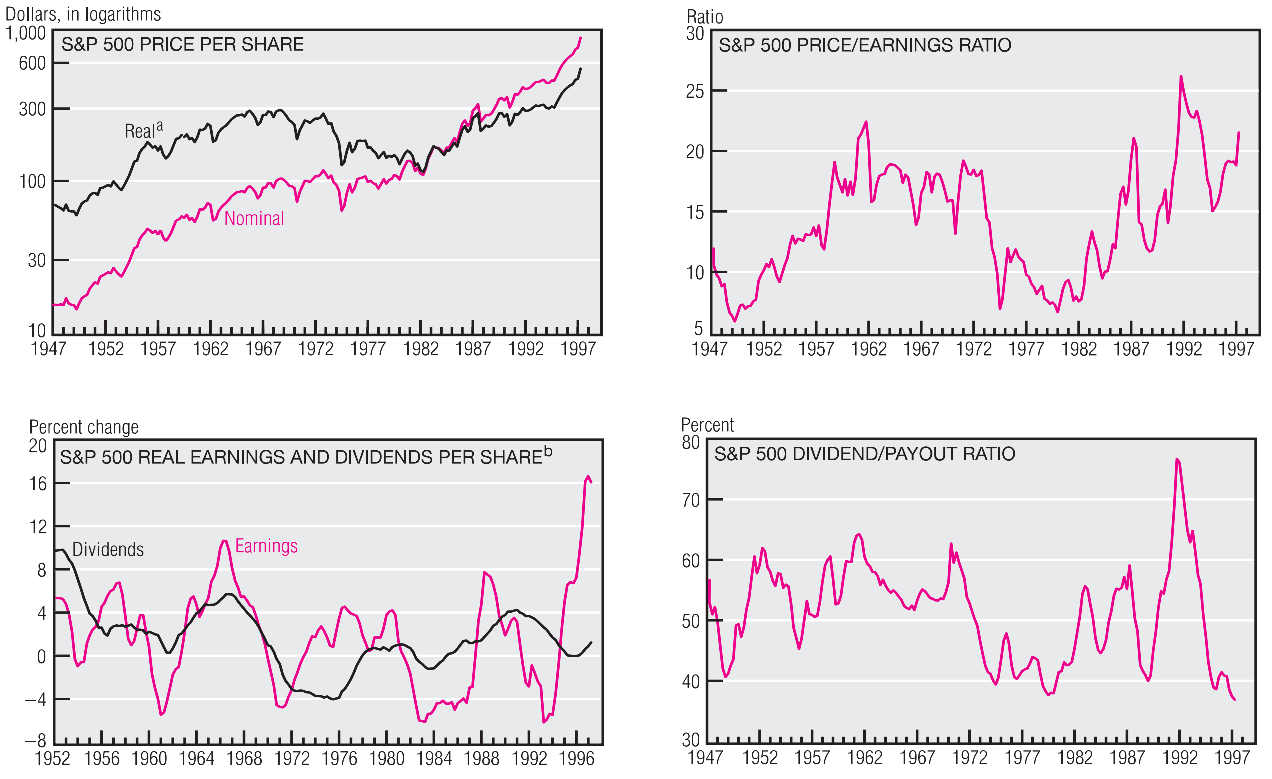
Introduction

In recent months, the meteoric ascent of the stock market has attracted intense interest. At the end of June 1997, the Standard and Poor's (S&P) 500 index stood more than 90 percent above its December 1994 level. Such an increase, however, is not unprecedented in the post–World War II era (see figure 1). Between September 1953 and September 1955—a period of just two years—the index also jumped more than 90 percent.

U.S. stock prices have been trending upward since 1982, although they have encountered many setbacks along the way. Nominal stock returns receive wider news coverage, but it is usually more meaningful to examine the data in real terms. The real S&P 500 has risen at a compounded annual rate of 26.8 percent since December 1994, 10.5 percent since October 1987, and 5.9 percent since 1975 (data are through June).

FIGURE 1

Stock Market Indicators



a. Deflated by the Consumer Price Index.
b. In 1983 dollars. Real earnings growth is the compounded growth rate of four-quarter total real earnings divided by four-quarter total real earnings four years earlier. Real dividend growth is the compounded growth rate of the current-quarter real dividend divided by the real dividend four years earlier.
SOURCES: Standard and Poor's Statistical Service, *Security Price Index Record*, various issues; and U.S. Department of Labor, Bureau of Labor Statistics.

Fundamentally, a stock's price is determined by the discounted value of its expected future dividends, which in turn derive from future earnings. When prospects for earnings growth are good, stock prices tend to rise. The price/earnings ratio (P/E)—the stock price divided by earnings per share—gives investors an idea of how much they are paying for a company's earning power. The higher the P/E, the more investors are paying, and hence the more earnings growth they are expecting. The average P/E of S&P 500 stocks has been rising over the past two years, approaching historic highs.

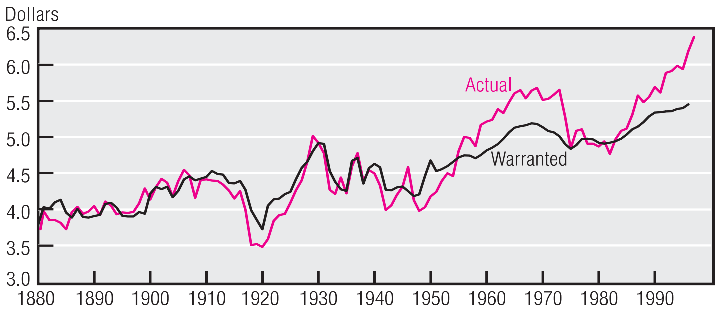
One clearly extraordinary fact associated with rising stock prices has been the phenomenal earnings growth over the past five years, which is viewed largely as a product of corporations' widespread efforts to cut costs and improve efficiency. Real earnings over this period increased at an average compounded annual rate of more than 12 percent. By contrast, dividend growth over the past five years has not been spectacular.

Much of the doubt about the recent stock market surge centers on whether it is justified by the fundamentals. Is it merely a speculative bubble? Or are stock prices correctly forecasting a healthy path for future dividends? If the surge is not justified by the fundamentals, then there is some risk of a precipitous decline in stock prices. Such situations concern monetary policymakers because they can create a sudden need to add liquidity to the financial system. This need for greater liquidity can complicate policymakers' efforts to pursue other objectives, especially if it conflicts with a need for anti-inflationary policy—a situation that occurred in October 1987.

To examine the causes of the stock market's recent ascent, we analyze the current relationship between stock prices, dividends, earnings, and returns. Our analysis reveals that no single

FIGURE 2

Actual vs. Warranted Real S&P 500 Index (Gordon Growth Valuation)



SOURCES: Standard and Poor's Statistical Service, *Security Price Index Record*, various issues; and U.S. Department of Labor, Bureau of Labor Statistics. Warranted series is based on authors' calculations.

fundamental element of standard stock valuation models can by itself explain the market's recent level. Rather, we conclude that stock prices could manifest both optimism about future dividend growth, which is predicated on the recent record growth in earnings, and a lower expected return, which reflects a diminished risk premium for holding equity. We also discuss the impossibility of knowing whether the implicit optimism is rational.[1]

The paper is organized as follows: Section I presents some common models of stock price variation that are based on a path of expected dividends and an expected rate of return. In section II, we examine alternative valuations of stock prices based on empirical models of the dividend process. We also suggest a modification of one such model that accounts for much of the recent rise in stock prices. The potential for lower expected returns is examined in section III. Section IV discusses the conditions that could accommodate higher expectations for dividends as well as lower expectations for returns. We offer concluding thoughts in section V.

I. Framework for Analysis

The standard framework for analyzing the valuation of capital stock is the *present-value model*.[2] In its basic form, a stock's price, P_t, is determined by the present value of its expected future dividends, D_{t+i}, and of the expected terminal price for the holding period K, P_{t+K}:

$$(1)\quad P_t = E_t\left[\sum_{i=1}^{K}\left(\frac{1}{1+R}\right)^i D_{t+i}\right] + E_t\left[\left(\frac{1}{1+R}\right)^K P_{t+K}\right],$$

where R is the expected return (or discount rate), assumed here to be constant.[3] As is common, we will assume that as K gets large, the expected present value of the terminal price shrinks to zero.

Under the conditions that dividends grow at a constant rate and the discount rate is time invariant, (1) simplifies to

$$(2)\quad P_t = \frac{(1+g)D_t}{R-g},$$

which is often expressed in log form

$$(3)\quad p_t = \log(1+g) + d_t + \log(R-g),$$

where lowercase p and d denote natural logarithms for prices and dividends. This relationship, traditionally called the Gordon growth model (Gordon [1962]), very compactly illustrates the connection between a stock's price, the current level of its dividend, the expected growth rate of dividends, and the discount rate. It is easy to see from (3) that the elasticity of price with respect to dividends is equal to one.

Because the derivation of (2) and (3) requires that both g and R be constant, the Gordon model does not lend itself to dynamic analyses that allow for time-varying discount rates or dividend growth rates. Nevertheless, this simple model can give insights into the fundamentals driving long-run fluctuations in the U.S. stock market. For example, the unitary elasticity of price with respect to dividends implies that swings in P_t should vary proportionally with swings in D_t. Assuming that $R-g$ equals 0.05, near its historical average, P_t will equal 20 times the current dividend.

Figure 2 illustrates the "warranted" value of the S&P 500 index, based on the Gordon valuation.[4] It shows that the data are broadly, though not precisely, consistent with this simple model

1 For a nontechnical discussion of some of the issues that we examine below, see Haubrich (1997).

2 This framework is sometimes referred to as the discounted-cash-flow model. For a lucid treatment of the relation between stock prices, dividends, and returns, see Campbell, Lo, and MacKinlay (1997).

3 When the expected return is constant, it is easy to show that (1) follows from the definition of *return*: $R_{t+1} = (P_{t+1} + D_{t+1})/P_t - 1$. Taking expectations of both sides of this relation and substituting $E_t R_{t+1} = R$ yields $(1+R)E_t P_t = E_t P_{t+1} + E_t D_{t+1}$, a linear difference equation that can be solved forward to obtain (1).

4 The historical series on dividends, earnings, and the S&P composite are described in Shiller (1989). Dividends are the totals for a year divided by the year's average producer price level. Stock prices are the real values for January.

over the past century or so. That is, low-frequency swings in S&P dividends are associated with long swings in the S&P index.

It is noteworthy, however, that stock prices fluctuate more widely than the warranted prices implied by Gordon's valuation rule, especially after 1950. The elasticities of long swings in stock prices with respect to dividends is about 1.5 for changes in the range of 10 to 30 years (see Barsky and De Long [1993]). That is, stock prices fluctuate about 50 percent more than the Gordon model implies. It is easy to see from (3) that stock prices are also extremely sensitive to changes in the discount rate and the projected dividend rate. Thus, a time-varying discount rate is also a candidate for explaining the excess variation.

Formal analysis of stock prices that allows the expected return to vary over time is much more difficult, since the relation between prices and returns is nonlinear.[5] To simplify the problem and make it analytically tractable, Campbell and Shiller (1988a, b) propose a log-linear approximation of the present-value framework that enables us to calculate asset price behavior under any model of expected returns, not just one for constant returns. Their formulation yields

$$(4) \qquad p_t = \frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \left[(1-\rho)d_{t+1+j} - r_{t+1+j}\right],$$

where k and ρ are "fixed" parameters defined in terms of the average log dividend/price ratio.[6]

Campbell, Lo, and MacKinlay (hereafter CLM [1997]) emphasize that equation (4) is a dynamic accounting identity. It illustrates clearly that if a stock price is high today, then it must be associated with higher future dividends, lower future returns, or some combination of the two.

Taking expectations of both sides demonstrates that (4) holds ex ante:

$$(5) \qquad p_t = \frac{k}{1-\rho} + E_t\left[\sum_{j=0}^{\infty} \rho^j \left[(1-\rho)d_{t+1+j} - r_{t+1+j}\right]\right].$$

Equation (5) is essentially a dynamic generalization of the Gordon model. It implies that high current stock prices must be associated with higher *expected* future dividends, lower *expected* future returns, or some combination of the two.

It is also easy to see from (5) that any transitory movements in expected future dividends or returns will have little impact on current prices. Persistent movements in these elements, however, can have substantial effects. Thus, in this framework, rational asset pricing implies that the reason for the stock market's recent rise must be the continuing substantial increase in the path of expected future dividends, the persistently lower expected returns of future years, or some combination of the two. Of course, if either set of expectations is not realized, then (5) guarantees that both sets will be wrong. That is, if the path of future dividends falls short of expectations, then ex post returns will be lower than anticipated. Similarly, if expected returns are lower than anticipated, then dividends will not be as high as expected.

The framework presented in this section does not give any guidance on the appropriate projections for future dividends or returns. It simply provides dynamic accounting identities, which follow from the definition of *return*. As identities, these relations can be used only to gauge the consistency of future expectations. To say anything about the particular path of stock prices requires a specification of the stochastic processes underlying the generation of the two fundamental series—expected future dividends and expected future returns.

II. Expected Future Dividends

The time series of log dividends is, to a rough approximation, a random walk (see Mankiw, Romer, and Shapiro [1985] and Kleidon [1986]). That is, changes in the level of dividends appear to be "permanent." Although the dividend series varies randomly over time, it has generally drifted upward. For example, real S&P 500 dividends increased an average of around 1.5 percent annually from 1871 to the present.

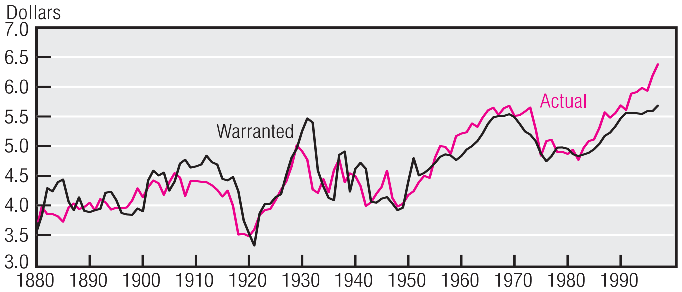
That dividends are approximated by a random walk accords well with a popular hypothesis of dividend formation—dividend smoothing. This theory holds that managers smooth dividends by setting dividend changes to reflect "permanent" changes in earnings. Such innovations are by their very nature unforecastable by investors, who are removed from management. Hence, both earnings and dividends tend to follow a random walk.

5 Of course, the problem is just as prevalent for time-varying dividends, as considered by Barsky and De Long (1993). Because they do not deal directly with this issue, they refer to their model as heuristic.

6 Campbell, Lo, and MacKinlay (1997) present evidence that the approximation misstates the average stock return but captures the dynamics of stock returns well, especially at monthly frequencies.

FIGURE 3

Actual vs. Warranted Real S&P 500 Index (Barsky–De Long Valuation)



NOTE: Warranted prices are based on the assumption that θ in equation (8) equals 0.96.
SOURCES: Standard and Poor's Statistical Service, *Security Price Index Record*, various issues; and U.S. Department of Labor, Bureau of Labor Statistics.

Most often it is assumed in the literature that dividend changes are stationary. Any innovation in the growth rate of dividends is thus temporary. That dividends are approximated by a random walk is consistent with the Gordon growth model and could explain the random-walk nature of "warranted" prices in figure 2.[7] As noted above, however, this approach fails to explain much of the stock price variability since 1950.

Barsky and De Long (1993)—hereafter BD—propose a modification in the Gordon growth model to explain some of this discrepancy. Specifically, they drop the assumption that the dividend growth rate has a constant mean known to agents throughout the sample. Instead, they postulate an environment in which investors estimate, period by period, a growth rate that is nonstationary and hence is itself a random walk. They redefine the Gordon valuation model as

$$(6) \qquad p_t = d_t + \log(R - g_t),$$

where g_t is an unknown "permanent" dividend growth rate.[8] They propose that g_t be treated as analogous to Milton Friedman's "permanent income" concept, which also is unknown, is changing over time, and must be reestimated every period.[9] For each future period, the expected future dividend growth rate is viewed as equal to some updated rate that is expected to persist indefinitely.

From (6), BD obtain conditions under which elasticity of price with respect to dividends is equal to the estimated value of 1.5. Maintaining the assumption that R is fixed, and using ∂ to denote a partial derivative, the elasticity of prices with respect to dividends is given by

$$(7) \qquad \frac{\partial p_t}{\partial d_t} = 1 + \left[\frac{1}{R - g_t}\right]\frac{\partial g_t}{\partial d_t}.$$

Since $R > g_t$, $\partial p_t / \partial d_t$ is greater than one *only if* the expected future growth rate is positively correlated with dividends. For purposes of illustration, assume that the $R - g_t$ term in the denominator of equation (7) is on the order of 0.05. To match the estimated elasticity, a 10 percent increase in the growth of dividends over a 20-year period (that is, 0.5 percent per year) would need to be associated with a shift in expected g_t of 0.25 percent. Thus, to account for the estimated relationship, about half of any shift in the average dividend growth rate over a 20-year period would be expected to persist indefinitely. Furthermore, the drift in the growth rate could be barely detectable and yet have a large impact on stock prices.

BD propose that agents estimate g_t using extrapolative forecasting methods. Such forecasts may be rational if the variable to be estimated is the sum of a random walk and a transitory white-noise error (Muth [1960]). This approach projects permanent dividend growth as a weighted geometric average of past dividend changes,

$$(8) \qquad g_t = (1 - \theta)\sum_{i=0}^{t} \theta^i \Delta d_{t-i} + \theta^t g_0,$$

where the weights decline geometrically with past values. Equation (8) thus implies that investors extrapolate past dividend growth into the future. Substituting (8) in (6) yields a series of warranted stock prices based on the BD modification of the Gordon model.[10]

Figure 3 illustrates such a series, on the assumption that θ equals 0.96.[11] Interestingly, in the period from 1880 to 1950, low-frequency

7 Stock prices are real values for January. Dividends are annual figures divided by the year's average producer price level. The historical series are printed in Shiller (1989).

8 BD implicitly assume that d_t applies to the period ahead and is known. Thus, they ignore second-order effects associated with estimating the first future dividend.

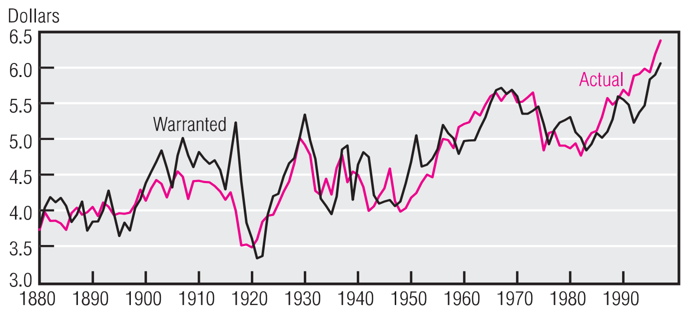
9 Although g_t varies over time, BD assume that the current g_t applies to all future periods.

10 BD note that (8) does not hold exactly in a stochastic model for the underlying dividend process because it does not allow prices today to be influenced by investors' knowledge that they will be revising their estimate of g in the future. For simplicity, they ignore such higher-order corrections.

11 BD calibrate this series for θ equal to 0.94 and 0.97. Although the latter value reduces the amplitude of warranted prices in the earlier period, the gap between warranted and actual prices is greater in the later period.

FIGURE 4

Actual vs. Warranted Real S&P 500 Index (Modified Barsky–De Long Valuation)



NOTE: Warranted prices are based on the assumption that θ in equation (8) equals 0.96.
SOURCES: Standard and Poor's Statistical Service, *Security Price Index Record*, various issues; and U.S. Department of Labor, Bureau of Labor Statistics.

swings in the warranted stock price are of greater amplitude than actual swings in the first part of the sample. After 1950, however, swings in the warranted stock price mimic longer-term swings in actual prices reasonably well—that is, until 1990. The latest bull market is not explained by this approach. Nevertheless, BD demonstrate that a significant proportion of the variation in stock prices can be explained by changes in expectations about future dividend growth.[12]

One of the important features of the recent bull market is that it has been associated with persistently strong earnings growth. As noted above, earnings growth over the past five years has exceeded that of any comparable five-year horizon in the post–World War II era.[13] Dividend growth, on the other hand, has been less than spectacular. What might account for this discrepancy? Kleidon (1986) questions whether the dividend-smoothing hypothesis adequately explains stock price movements:

> [*The problem of dividend smoothing*] *has important implications for all research that attempts to infer the properties of an infinite stream of future dividends from some finite ex post set of dividends that are under some control of management. Empirical evidence suggests that management takes care to create a smooth short-run dividend series that may not reflect one for one the fortunes of the firm as determined primarily by its earnings and investment opportunities. Ceteris paribus, the less variable the dividend stream, the more variable will be the price series that comprises the present value of future dividends. For example, a firm seeking to finance expansion internally may withhold all dividends over some finite period, with an implicit promise of some future (perhaps liquidating) dividends.* (p. 975)

The central point is that corporate management may see great investment opportunities, which they may choose to finance with retained earnings. In this situation, dividends would be less than otherwise. Hence, lower dividends would be incorrectly signaling lower future earnings. The fact that dividend growth is much smoother than earnings could reflect the fact that periods of persistently high earnings growth are also associated with increased internal financing and hence slower dividend growth.

Internal financing is not the only potential problem for the dividend-smoothing hypothesis. In the late 1980s, corporations began to repurchase shares on a large scale (CLM [1997], p. 287). Such strategies have the same "distorting" effect that internal financing has on the prospective information content of ex post dividends. Reports of record share repurchases over the past year suggest that this phenomenon may explain part of the recent discrepancy between earnings and dividend growth.[14]

These problems suggest that "permanent" changes in earnings may be better estimated directly from the series of ex post earnings. To assess this hypothesis, we apply the extrapolative forecasting methods proposed by BD for the earnings series to estimate the "permanent" growth rate of earnings, that is, the growth rate expected to persist indefinitely. If the dividend payout ratio has a fixed mean, then over long horizons, dividends will grow at the same rate as earnings.

Figure 4 compares a warranted price series based on equation (6) and a "permanent" growth rate of dividends based on equation (8) with log earnings in place of log dividends. As

12 Donaldson and Kamstra (1996) propose a dividend-forecasting approach based on a nonlinear ARMA-ARCH-Artificial Neural Network model. The present value of out-of-sample dividend forecasts from their model yields fundamental prices that reproduce the magnitude, timing, and time-series behavior of the boom and crash in 1929 stock prices. They do not, however, apply their model to recent history.

13 In the five-year periods ending 1899, 1925, and 1926, earnings growth exceeded that of the recent period. However, these episodes followed persistent periods of earnings declines.

14 For example, Barrett (1996) reports that in the first 10 months of 1996, 1,185 companies announced that they intended to buy back a portion of their publicly traded securities, repurchasing shares worth $129 billion. The report called this volume a record pace, citing figures from Securities Data Corp.

in the BD valuation, the warranted series is much more volatile in the early part of the sample. The series based on the direct measure of permanent earnings, however, explains a greater part of the recent ascent in stock prices, but fails to account for all of it. Indeed, as of June 1997, the actual index level had increased even more than the warranted one.[15]

Some readers may be surprised that such simple models can explain so much of the long-horizon swings in stock prices, but these approaches get a lot of mileage out of expected dividends as the primary mover of the stock market. Like BD, we have postulated that shocks to the dividend growth rate include a small permanent element. We propose, however, that this element can be estimated directly from the history of earnings growth. Investors rationally extrapolate (in the sense of Muth [1960]) past earnings growth into a future dividend growth rate.

If it depended on the existence of a large unit root in the earnings process, this approach might seem ludicrous. It would defy evidence that the U.S. dividend process is reasonably well approximated by a random walk with constant drift. We choose, however, a value of θ equal to 0.96 for generating the extrapolative forecasts used in figures 3 and 4. Such a parameter value implies a unit root in the dividend process that creates only a very small share of annual dividend growth volatility. Moreover, BD perform a Monte Carlo simulation demonstrating that the sample size and the magnitude of permanent growth rate shocks are both too small to be informative on the value of θ. The data just do not refute the hypothesis that dividend growth has a small permanent component.

The examples we have discussed in this section, however, explicitly assume that the expected return is time invariant, an assumption that has lately become untenable. A number of empirical studies in recent years have demonstrated that stock returns are predictable using information other than past returns.[16] The predictability of time-varying returns implies that rational investors' expectations about returns vary.

III. Expected Future Returns

It is well known that the stock market has generally been kind to those who buy and hold a diversified portfolio throughout their working lifetimes. Returns to holding stock have been remarkably stable over very long horizons. For example, Siegel (1994) finds that during three periods—1802–70, 1871–1925, and 1926–90—real compounded equity returns were 5.7, 6.6, and 6.4 percent. Such consistency suggests that expected returns over very long horizons might also be quite stable.

Ex post equity returns over horizons of 10 years, however, paint a somewhat different picture. For example, between 1881 and 1997, 10-year returns on S&P stocks varied between 23.1 and –3.7 percent. Ex post returns, however, are not the same as expected returns, which are unobservable. Thus, evidence concerning the variability of expected returns is necessarily indirect, based on the forecastable component of returns.

Estimates using different information sets suggest that changes in expected returns can be substantial. Blanchard (1993), for example, estimates a series of long-term expected returns of New York Stock Exchange companies from 1927 to 1992. Using the dynamic version of the Gordon valuation model, he finds that the expected real return ranged between 10 percent in 1950 and about 2.5 percent in 1970.[17]

From estimates of both expected stock returns and expected bond yields, Blanchard extracts an estimate of the equity premium—the difference between the return on stocks and the yield on bonds—over time. Because stocks have historically been riskier than bonds, investors generally expect a higher return for stocks. Blanchard estimates that the trend in the equity premium began drifting down after 1950. This result suggests that there may in fact be a permanent reduction in expected future stock returns.[18]

Blanchard notes that the declining trend in the equity premium accords with the increasing importance of institutional investors since 1950. For example, the share of equities held by state, local, and private pension funds ballooned from 1 percent in 1950 to 9 percent in

15 When this article went to press on August 13, earnings were available only for the first quarter of 1997. We estimated the warranted price on the basis of these data.

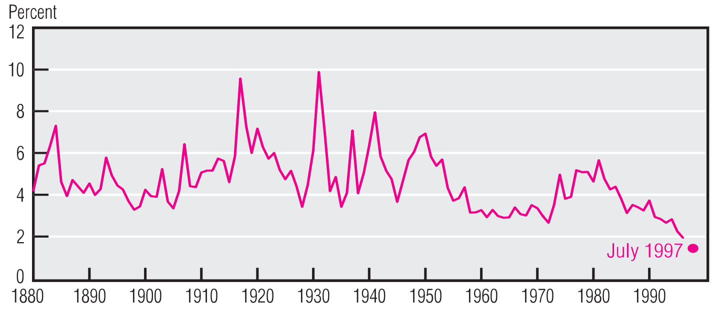
16 See CLM (1997, chapter 7) for a review of this evidence.

17 Blanchard essentially backs out an estimate by adding projections of the expected long-term growth rate, measured as the annuity value of the growth rate of future dividends.

18 A lower equity premium would solve a puzzle posed by Mehra and Prescott (1985). They note that it is difficult to reconcile the post–World War II equity premium within the standard utility-based models of asset pricing. Specifically, using a model with constant relative risk aversion, they show that the realized equity premium of the postwar era implies a level of risk aversion too high to be consistent with their model.

FIGURE 5

S&P 500 Index Dividend/Price Ratio



SOURCE: Standard and Poor's Statistical Service, *Security Price Index Record*, various issues.

1970 to 29 percent in 1993. Blanchard argues that pension managers have a mandate to think in terms of longer horizons and to take advantage of an attractive equity premium. This, of course, implies that private investors generally have not been farsighted.

Benartzi and Thaler (1995), however, propose a theory of behavior under uncertainty that explains the large equity premium. Their approach is based on the notion of loss aversion developed by Kahneman and Tversky (1979), with preferences defined over gains and losses rather than consumption, and losses given greater weight than gains.[19] Benartzi and Thaler show that loss aversion over short horizons can rationalize investors' reluctance to hold stocks even though they are aware of a large premium. If such behavior dominated the pricing of stocks before pension funds proliferated, it could explain why the equity premium was so high and why it may have become permanently lower.

Blanchard's estimate of a declining equity premium—and hence a declining expected rate of return on stocks—is consistent with the obvious trend change in the relationship between dividends and stock prices, as implied by the simple Gordon growth approach. For example, figure 2 illustrates that since about 1950, stock prices have persistently exceeded the warranted prices implied by the Gordon valuation. Such a discrepancy diminishes if one allows for a declining required rate of return.

■ **19** The fundamental choice paradigm of this approach is supported by experimental evidence.

IV. Discussion

We have seen that a substantial part of the stock market's rise can be rationalized either as an expectation that dividend growth will surpass its historical norm indefinitely, or that the equity premium has so diminished that expected future returns are substantially lower, or both. How different must things be in the future to be reconciled with recent stock market levels? To get some perspective on the potential for each of the two fundamental elements, it is useful to consider the dividend/price ratio.

In late July 1997, the S&P 500 index stood near the 950 level. Dividends per share over the previous year were about $15. The dividend/price ratio was thus around 1.5, a historical low (see figure 5). In terms of the simple Gordon model (equation [2]), the dividend/price ratio equals

$$D_t/P_t = (R - g)/(1 + g).$$

Starting from this relationship, one can examine the consistency of alternative pairs of assumptions about R and g. There are, of course, an infinite number of combinations of these two variables consistent with a dividend/price ratio equal to its recent level. Based on the above analysis, consider the following example: If one believes that future expected returns on stock are now at 4 percent, the Gordon model implies that the expected long-term dividend growth rate would need to be about 2½ percent (about one percentage point higher than its historical average). Although this may seem reasonable, it is clearly a substantial shift if placed in a historical context.

Could such optimism be a misguided consequence of the relative tranquility the U.S. economy has experienced over the last 15 years? Could extrapolation induce complacency about the future? Are risk premiums too low? If the current valuation of the stock market required dividend growth over the next 125 years to be 1 percent higher than it was over the previous 125 years, then markets might seem unduly optimistic.

The dynamic version of the Gordon model, on the other hand, suggests the possibility that persistent, extraordinary events can drive things in the short run. If this is so, could the market be forecasting a sharp increase in dividend growth for a sustained period before settling back to its long-term rate? Given the record growth of earnings recently, it is conceivable that dividends could accelerate sharply for several years and persist at a higher growth rate before decelerating to a more normal clip. This would leave dividends at a higher level from which to grow.

To get a sense of the potential effect of a short-run surge in dividend growth, it is useful to consider a hypothetical comparison. As a benchmark, assume an initial dividend of $1 per share, expected dividends growth at a constant rate of 2 percent, and a constant discount rate of 6 percent. According to the simple Gordon model (equation [2]), the present value of the expected dividend stream would equal $25.50. As an alternative, consider a dividend stream that is expected to grow 6 percent annually for five years and then return to a normal growth rate of 2 percent. The present value of dividends over the first five years is $5 (the first term in equation [1]).[20] The present value of the dividend stream beginning in the sixth year is $25.50.[21] Thus, the present value of the total stream is $30.50, almost 20 percent higher than the benchmark. If the 6 percent dividend growth rate were expected to persist for 10 years before returning to normal, the alternative variation would be almost 40 percent higher.

It should be stressed here that, historically speaking, dividends are reasonably well approximated by a random walk. Thus, our example attributes to investors a forecasting skill that is not evident in the financial literature. Perhaps investors are anticipating near-term rewards because of two heavily publicized events. First, much has been made of corporate restructuring. Increased focus on shareholders' interests reportedly has equipped corporations to respond better to market incentives. Hence, it is argued that they should be much more profitable in the future than in recent decades.

Second, the information revolution has led to the rapid development of new technologies, which are only now beginning to be realized. Greenwood and Yorukoglu (1997), for instance, present a model in which technological innovation is embodied in new producer durables or services. Their analysis suggests that the U.S. economy is at the dawn of an industrial revolution as significant as that associated with the development of the steam engine. The optimism implied by such hypotheses is difficult to justify because it is based on low-frequency events for which there is limited empirical evidence.

Alternatively, it is conceivable that the recent surge in stock prices reflects a transitory but persistent decline in expected returns. To assess the potential for this explanation, it is useful to consider an example given by CLM (1997, p. 265). Suppose expected returns are described by the following AR(1) process:

(9) $$E_t[r_{t+1}] = r + x_t,$$

where r is a constant and x_t is described by

$$x_{t+1} = \phi x_t + \xi_{t+1} \qquad -1 < \phi < 1.$$

When ϕ is close to one, the process is typically described as highly persistent. The variance of x, σ_x^2, is related to the variance of the innovation ξ_t, σ_ξ^2, by $\sigma_\xi^2 = (1 - \phi^2)\sigma_x^2$. The implication of x_t on the stock price is obtained by substituting (9) into the expected-return component of the dynamic Gordon growth model:

(10) $$E_t\left[\sum_{j=0}^{\infty} \rho^j r_{t+1+j}\right] = \frac{r}{1-\rho} + \frac{x_t}{1-\rho\phi}.$$

The last term gives the effect on the stock price of the expected return's variation through time. To illustrate, consider a persistent process where $\phi = 0.9$. Since ρ is equal to about 0.96, a 1 percent increase in the expected return today would reduce the stock price by about 7.3 percent. If $\phi = 0.5$, a 1 percent increase in the expected return would reduce the stock price by 1.9 percent. This illustrates that relatively small changes in expected returns can have large impacts on stock prices, if such changes persist.

Studies show that the dividend/price ratio forecasts future returns, especially over horizons of four years (see, for example, CLM [1997], pp. 267–70). Evidence indicates that a low dividend/price ratio is associated with lower future returns. Figure 5 suggests that returns are likely to be below normal over the next few years.

We have thus far assessed alternative assumptions independently. Cochrane (1994) examines both dividends and returns in a two-variable VAR, including the dividend/price ratio as an explanatory variable in both equations. He finds that shocks to dividends have immediate, permanent effects on both dividends and stock returns. Shocks to returns, holding dividends constant, have strictly transitory effects on returns, but no effects on dividends. Moreover, the dividend/price ratio forecasts returns much more strongly than dividends alone.

20 In this case, the growth factor and the discount rate cancel in each period.

21 Again, because expected dividends grew at a rate equal to the discount factor, the current present value of the future stream beginning in the sixth year equals the benchmark value.

Thus, he argues, returns rather than dividends adjust to bring the ratio back to its mean.[22]

Cochrane's results do not allow for a permanent reduction in the expected rate of return. Had the equity premium been permanently reduced, one might expect a permanently lower dividend/price ratio, which is assumed to be stationary in the Cochrane analysis. But, how low could the expected return go before it would be unduly optimistic? Is a 4 percent expected rate of return too low to be associated with the risk inherent in the stock market? The answer to that question is beyond the scope of this paper.[23]

V. Concluding Thoughts

Our analysis shows that if priced rationally, the stock market's current level must imply one of three things: either investors expect dividends to accelerate and persist at higher levels for some substantial period, or investors are expecting much lower returns than the historical norm of around 6 percent in real terms, or some combination of these prevails. Of course, such expectations may not be realized, and investors betting on an acceleration of dividends would ultimately be disappointed in their returns if such an acceleration failed to materialize.

The U.S. economy has enjoyed a long, healthy expansion in the 1990s. Inflation has been contained, and there is little evidence of any imbalances to suggest that the end of the expansion is imminent. Earnings growth during this period has been extraordinary. If investors expect a small component of the recent surge in earnings to persist indefinitely, then stock prices would be higher than traditional valuation approaches indicate. Our analysis, however, suggests that earnings growth is not the whole story.

The development of pension funds and their mandate to maintain a focus on a long horizon may have led to a lower equity premium. If pension fund managers generally believe that a 4 percent return is sufficient reward for the risk, given market rates on alternative instruments, then expectations on dividend growth need not be so optimistic. A 4 percent real return is probably not inconsistent with standard economic models of behavior under uncertainty and the level of risk associated with stock prices. If investors are expecting higher returns, they are likely to be disappointed unless dividend growth accelerates substantially. Whether such unrealized expectations would be irrationally exuberant, we cannot know.

■ **22** Cogley (1996) finds that the quarterly standard deviation of the nominal discount factor must be at least 9.4 percent. He concludes that a model which implicitly assumes expected returns to be constant is not likely to describe the market very well.

■ **23** On the other hand, standard utility analysis suggests that the historical level of the equity premium is too high to be consistent with "rational" economic behavior.

References

Barrett, K. "Record Year for Buybacks." Cable News Network, *Financial Network*, October 28, 1996. http://www.cnnfn.com/hotstories/companies/9610/28/stockbuybacks/.

Barsky, R.B., and J.B. De Long. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, vol. 108, no. 2 (May 1993), pp. 291–311.

Benartzi, S., and R.H. Thaler. "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, vol. 110, no. 1 (February 1995), pp. 73–92.

Blanchard, O.J. "Movements in the Equity Premium," *Brookings Papers on Economic Activity*, no. 2 (1993), pp. 75–138.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. *The Econometrics of Financial Markets.* Princeton, N.J.: Princeton University Press, 1997.

Campbell, J.Y., and R.J. Shiller. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, vol. 1, no. 3 (Fall 1988a), pp. 195–228.

———, and ———. "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, vol. 43, no. 3 (July 1988b), pp. 661–76.

Cochrane, J.H. "Permanent and Transitory Components of GNP and Stock Prices," *Quarterly Journal of Economics*, vol. 109, no. 1 (February 1994), pp. 241–65.

Cogley, T. "Why Do Stock Prices Sometimes Fall in Response to Good Economic News?" Federal Reserve Bank of San Francisco, *Economic Letter*, No. 96–36, December 13, 1996.

Donaldson, R.G., and M. Kamstra. "A New Dividend Forecasting Procedure that Rejects Bubbles in Asset Prices: The Case of 1929's Stock Crash," *Review of Financial Studies*, vol. 9, no. 2 (Summer 1996), pp. 333–83.

Gordon, M. *The Investment, Financing, and Valuation of the Corporation*. Homewood, Ill.: R.D. Irwin, 1962.

Greenwood, J., and M. Yorukoglu. "1974," *Carnegie-Rochester Conference Series on Public Policy*. New York: Elsevier, 1997 (forthcoming).

Haubrich, J.G. "Stock Market Fundamentals," Federal Reserve Bank of Cleveland, *Economic Commentary*, January 15, 1997.

Kahneman, D., and A. Tversky. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, vol. 47, no. 2 (March 1979), pp. 263–91.

Kleidon, A.W. "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, vol. 94, no. 5 (October 1986), pp. 953–1001.

Mankiw, G.N., D. Romer, and M.D. Shapiro. "An Unbiased Reexamination of Stock Market Volatility," *Journal of Finance*, vol. 40, no. 3 (July 1985), pp. 677–87.

Mehra, R., and E.C. Prescott. "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, vol. 15, no. 2 (March 1985), pp. 145–61.

Muth, J. "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, vol. 55, no. 290 (June 1960), pp. 299–306.

Shiller, R.J. *Market Volatility*. Cambridge, Mass.: MIT Press, 1989.

Siegel, J.J. *Stocks for the Long Run: A Guide to Selecting Markets for Long-Term Growth*. Burr Ridge, Ill.: Irwin Professional Publishing, 1994.

Workforce Composition and Earnings Inequality

by Mark E. Schweitzer

Mark E. Schweitzer is an economist at the Federal Reserve Bank of Cleveland. The author gratefully acknowledges the advice of Trudy Cameron, Ben Craig, Bruce Fallick, David Lewin, Dan Mitchell, and Finis Welch. He also thanks workshop participants at UCLA, Texas A&M, Case Western Reserve University, and the Federal Reserve Banks of Cleveland and Dallas.

Introduction

As the United States passed through another election phase last fall, we again heard about the increasingly unequal earnings prospects of America's workforce. While there are many suggested remedies, making educational opportunities more available to all is the most common.[1] Proponents of this approach believe that rising returns to education can be attenuated by increasing the supply of highly educated workers and reducing the supply of less skilled workers. This follows from analyses indicating that education is the primary factor contributing to earnings inequality.[2]

Existing research in this area has typically focused on a single demographic group rather than on how demographic groups' earnings relate.[3] Juhn and Murphy (1997) extend their earlier analysis on white males to both sexes by considering the effects of marriage and family structure on family inequality. They find that like workers tend to marry one another, increasing the earnings gap between families. This approach returns the focus to general workforce inequality, but includes four major demographic groups in a generalized inequality decomposition based on estimating their human capital returns independently. The most notable results of this analysis that could not be ascertained in previous research are 1) the increasing share of women in the workforce and their increasing realized tenures have reduced earnings inequality, and 2) a larger portion of the variation in earnings is associated with the changing composition of the workforce, rather than with changing returns to human capital investments.

The three main qualitative results of the existing research are confirmed in Juhn and Murphy's study, although the levels of these factors are altered somewhat by either the different population or the procedures used in their analysis: 1) educational differences are the primary variable associated with rising inequality, 2) industry affiliation and wage differentials are associated with rising inequality, and 3) wage

1 A good summary of these proposals can be found in Freeman (1996) and in the responses to his article.

2 Juhn, Murphy, and Pierce (1993) and Murphy and Welch (1992) make a strong case for focusing on education. For a broader survey of the literature, see Levy and Murnane (1992).

3 Karoly (1992) studies the importance of gender and other factors individually, while Burtless (1990) presents an extensive comparison of wage inequality between men and women.

differences by experience level (age) or region have little impact on inequality. What is new is that Juhn and Murphy's results apply to the entire full-time/full-year workforce, even though that population has changed dramatically (particularly through the addition of more women and minorities and the higher educational levels attained, especially by those two groups). This suggests that typical policy remedies may take a long time to overcome the trend toward increasing inequality.

While the widening disparity in Americans' earnings is a heavily cited and discussed phenomenon, pinpointing its source is a complex exercise. The level of earnings inequality in a society is determined by interactions among many factors. For example, one worker's higher education level may be offset by another's greater experience, yielding no inequality between them. The potential for interactions among earnings factors is large in a diverse workforce, because diversity introduces many potentially offsetting and augmenting sources of inequality. In fact, as previously documented, the wage structure in the United States has been altered along several dimensions over the last two decades: Educational differentials have expanded, experience profiles have steepened, women's wages have drawn closer to men's, and so on.

The potential for interactions among these factors is not merely of academic interest. Such interactions may alter the impact of wage structure changes, including those encouraged by public policy. Again, it is instructive to look at an example. Increasing the educational level of the highest-paid members of a demographic group may boost the earnings disparity within that group while clearly lowering the societal level of inequality. Ultimately, the effect of any change in the earnings structure on earnings inequality depends on the covariation of the altered factors with the other earnings characteristics of the studied population.

To account for interactions among earnings factors, this paper applies a generalized decomposition to Current Population Survey data on the earnings of all full-time/full-year U.S. labor force participants. The decomposition is implied by a model of earnings that encompasses a broad set of variables *simultaneously* in order to describe sources of earnings: education, experience, industry, and region. Furthermore, to better account for the changing composition of the workforce, the model is estimated independently for each of four race/sex groups (minority males, minority females, white males, and white females). This allows distinct wage determination patterns to emerge for these groups, which might alter the covariation of wages between them.

The remainder of the paper is organized as follows. Section I lays out a framework for determining earnings factors in a diverse workforce. Section II decomposes earnings inequality within this framework and extends the framework to consider the role of rising returns versus demographic changes. Section III summarizes and reconsiders policy prescriptions for the increasing earnings gap.[4]

I. Inequality Implications of Earnings Models

The treatment of earnings inequality in this paper follows the approach of Mincer's (1958) seminal work on human capital and the distribution of personal income—a specification that is now typically used in predictive models of earnings. Mincer used his model to stress that inequality due to human capital differences, a fundamental source of earnings inequality, should be separated from other sources of disparity. The result of differences in human capital investment can be summarized by the classic earnings equation, developed in Mincer (1974):

$$(1) \quad \ln W_i = \ln W_{0i} + rS_i + b_1X_i + b_2X_i^2 + v_i,$$

where $\ln W_{0i}$ is the wage for a worker's innate ability, S_i is years of schooling, X_i is years of experience, and v_i includes unobserved individual differences.

Equation (1) is extended below to provide a better fit with the actual experience profile, as suggested by Murphy and Welch (1990). An important extension of Mincer's framework is to allow workers to gain returns for working in their current industry. This is the logical extension of job-specific human capital (Oi [1962]) to industries. The final factor typically included in earnings models (other than race and sex, which receive a more careful treatment below) is the location of an individual's residence.[5] A simple but limiting means of accounting for these effects is to assume that the differences are constant across characteristics. Then, the earnings equation becomes

■ **4** Construction of the data set, which largely follows the "committed worker" restrictions of Juhn, Murphy, and Pierce (1993), is described in the appendix.

■ **5** See Eberts (1989) for a detailed look at regional wage differences.

TABLE 1

Race/Sex-Group Relative Wages and Workforce Shares (percent)

	1972	1978	1984	1990
Level	**Estimated Value of Race/ Sex-Group Differentials**			
White female	–36.88	–33.49	–29.10	–24.69
Minority female	–40.76	–35.84	–32.46	–28.25
Minority male	–20.71	–15.75	–16.87	–13.72
Frequency	**Percentage of Full-Time/ Full-Year Workforce**			
White male	61.20	56.76	53.27	51.15
White female	28.58	31.72	34.24	35.77
Minority female	4.23	5.17	6.10	6.45
Minority male	5.99	6.34	6.39	6.63

NOTE: Percentages are in terms of weekly wages evaluated around the intercept.
SOURCE: Author's calculations.

$$(2)\quad \ln W_i = \ln W_{0i} + b_1 S_i + b_2 X_i + b_3 D_i^{ind} + b_4 D_i^{oth} + v_i,$$

where S_i represents a vector of schooling-level indicators, X_i is a vector of quadratic experience terms, D_i^{ind} represents industry-specific effects, and D_i^{oth} represents regional effects. In the estimation, the rates of return for the earnings factors are allowed to change from year to year. Thus, the value and distribution of these skills and other factors are allowed to vary with shifts in labor supply and demand.

Accounting for Race and Sex Differences

Why account for changes in the racial and sexual composition of the workforce? Since earnings data were first collected, systematic differences in demographic groups' wages have been apparent. Between 1972 and 1990, a large shift occurred in the demographic composition of the full-time/full-year workforce. Table 1 illustrates both of these trends. The estimated differentials represent the coefficients for dummy variables in an earnings equation as specified above, that is, one which controls for experience, education, aggregated industries, and regions. The value of race/sex-group differentials falls by approximately one-third for each of these groups over the 18-year period, while the relative role in the full-time/full-year labor force for all four groups is rising.

Although the specification developed in equation (2) is a standard framework for measuring wage differences, particularly after including race and sex dummy variables, it provides little information about either the sources of race/sex differences or their effects on overall inequality. In particular, if returns to measured skills vary systematically by race/sex groups, then as the composition of the workforce changes, the estimated rates of return would be altered without any variation in the underlying rates of return for specific race/sex groups. A flexible specification that accounts for these differences by allowing complete variation in rates of return for all factors and for the error term by race/sex group is described in equation (3):

$$(3)\quad \ln W_i = \sum_{C \in \{\text{race/sex groups}\}} (\ln W_{0C} + b_{1C} S_i + b_{2C} X_i + b_{3C} D_i^{ind} + b_{4C} D_i^{reg} + v_{iC}),$$

where C indicates the race/sex group of individual i.

Returns to factors could vary by race or sex for several reasons. Returns to observed factors could differ because of qualities unobserved by the econometrician but seen by market participants. Alternatively, race or sex discrimination could be limited to particular sections of the labor market or restricted to certain factors. One clear source of differences in rates of return by race/sex is the variation in actual experience for given levels of potential experience observed in the Current Population Survey. Differing rates of return could also develop as a response to workers' inability to unbundle their set of skills, as shown by Heckman and Scheinkman (1987). They prove that differences in rates of return for observed and unobserved skill factors can vary by group if the proportions of skills vary and workers cannot market their skills separately.

Covariance Structure of the Model

The implications of this model for mean earnings are well known; however, its implications for earnings *inequality* have been applied only infrequently in the recent surge of inequality

TABLE 2

Correlations between Regression Components

1972	Experience	Education	Industry	Region
Experience	1.0 (0.00)			
Education	–0.1738 (0.0001)	1.0 (0.00)		
Industry	0.0721 (0.0001)	–0.0894 (0.0001)	1.0 (0.00)	
Region	0.0020 (0.6947)	0.0405 (0.0001)	0.0527 (0.0001)	1.0 (0.00)
1990	**Experience**	**Education**	**Industry**	**Region**
Experience	1.0 (0.00)			
Education	–0.1595 (0.0001)	1.0 (0.00)		
Industry	0.0928 (0.0001)	–0.1204 (0.0001)	1.0 (0.00)	
Region	0.0051 (0.2574)	0.0360 (0.0001)	0.0234 (0.0001)	1.0 (0.00)

NOTE: Figures in parentheses are probability values for the null hypothesis that the correlations are zero.
SOURCE: Author's calculations.

literature.[6] Consider a scenario of increasing returns to a single factor—education, for example. The standard decomposition of inequality by subgroups provides a simple comparison of mean earnings by industry and concludes that inequality rises. In terms of equation (2), one treatment of this hypothesis is that the range of vector b_1 is increased, as measured by $\max(b_1) - \min(b_1)$. This raises the variance of the term b_1S_i, but the effect of increasing the range of b_1 on the variance of earnings also depends on the signs of the covariances.

$$(4)\quad \frac{\partial \operatorname{var}(\ln W_i)}{\partial \operatorname{range}(b_1)} = \frac{\partial \operatorname{var}(b_1 S_i)}{\partial \operatorname{range}(b_1)} + 2\frac{\partial \operatorname{cov}(b_1 S_i, b_2 X_i)}{\partial \operatorname{range}(b_1)} + \ldots + 2\frac{\partial \operatorname{cov}(b_1 S_i, v_i)}{\partial \operatorname{range}(b_1)}.$$

Only the first and last terms of this derivative may be signed: The first is unambiguously positive, and the last (the covariance with the error term) is always zero by ordinary least squares.

Empirically, these covariances are a substantial and statistically significant portion of total wage variation, as indicated by the correlations in table 2. The reported correlations are for a regression of individual log wages on four of the variable categories discussed throughout this paper: experience, education, industry, and region. In addition to being generally significant, these correlations may change over time, as a cursory comparison of the 1972 and 1990 results indicates. Individual returns to education appear to be especially correlated with two other recognized earnings factors: experience and industry. This is not surprising, since educational levels are higher for younger cohorts, and education is clearly associated with one's industry choice.

Neglecting the covariances among explanatory variables affects the interpretation of the impact of industry wage differentials on earnings inequality. For example, Freeman (1991) argues that the loss of labor union premiums for low-skilled workers has exacerbated U.S. earnings inequality. Standard subgroup decompositions would be inappropriate without including other observed determinants of industry wage differentials, since they would indicate only the effect of union wage differentials. Freeman's point is that inequality is lower because of a negative covariance between union effects and skill factors.[7] An inequality decomposition should account for this negative covariance, thereby reducing the earnings inequality associated with union wage differentials. Without direct observation of union status, union effects can be viewed as a component of $b_3D_i^{ind}$, and the argument applies to industry premiums as well. Accounting for covariances can be similarly justified for most factors considered in the earnings inequality literature.

Equation (3) shows the more complicated covariance structure to be summarized by the decomposition. This extension alters the interpretation of the factors and allows for comparisons across groups. A change in the rate of return a single group is paid for a factor depends on both the covariance structure with that group's other factors and the covariances between that group's and other groups'

6 Smith and Welch (1979) did recognize the importance of covariances between explanatory variables in their analysis of race differences in earnings inequality. A similar technique is applied to prime-age white males in Blackburn (1990).

7 Freeman avoids this criticism by not performing an explicit inequality decomposition. Instead, he applies shift/share analysis to regression estimates after controlling for education.

TABLE 3

Education Differentials by Race/Sex Group (percent)

	Estimated Value of Education Differentials			
1972	**White Men**	**Minority Men**	**White Women**	**Minority Women**
High school dropout	–18.22	–15.99	–11.19	–13.28
Some college	14.63	7.29	12.48	9.15
College graduate	50.54	44.52	51.71	72.20
Post-graduate	75.49	77.01	80.16	124.49
1990	**White Men**	**Minority Men**	**White Women**	**Minority Women**
High school dropout	–25.72	–15.42	–21.42	–15.06
Some college	19.01	20.12	18.26	15.47
College graduate	58.69	51.69	65.02	52.29
Post-graduate	89.47	99.58	102.63	102.20

NOTE: Percentages are in terms of weekly wages evaluated around the race/sex-group intercept.
SOURCE: Author's calculations.

FIGURE 1

Experience–Earnings Profiles: 1972

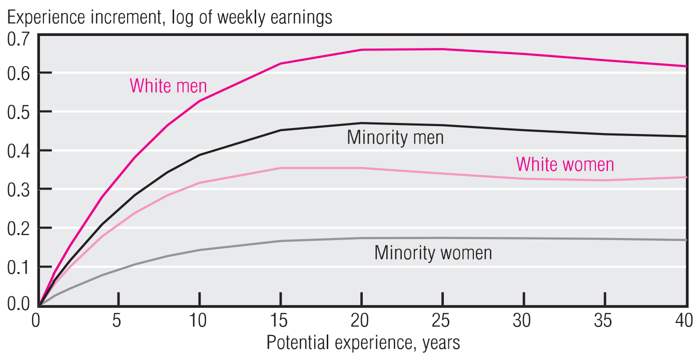
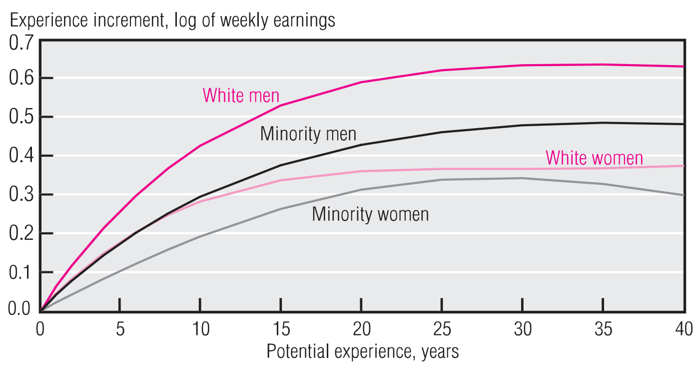


FIGURE 2

Experience–Earnings Profiles: 1990



SOURCE: Author's calculations.

wages. Repeating the earlier example of an increase in the rate of return to education for workers in group 1 of four race/sex groups, we have

$$(5) \quad \frac{\partial \operatorname{var}(\ln W_i)}{\partial \operatorname{range}(b_{1,1})} = \frac{\partial \operatorname{var}(b_{1,1}S_i)}{\partial \operatorname{range}(b_{1,1})} + 2\,\frac{\partial \operatorname{cov}(b_{1,1}S_i,\, b_{2,1}X_i)}{\partial \operatorname{range}(b_{1,1})} + \ldots + 2\,\frac{\partial \operatorname{cov}(b_{1,1}S_i,\, v_i)}{\partial \operatorname{range}(b_{1,1})} + 2 \sum_{j=2,3,4} \frac{\operatorname{cov}(b_{1,1}S_i, \ln W_{i\in j})}{\operatorname{range}(b_{1,1})} \cdot$$

Equation (5) raises the possibility that certain factors could increase earnings inequality within a group and yet reduce populationwide inequality.

Race/sex groups have different returns to factors, and these returns change across the period. Educational differentials are shown in table 3. With the exception of highly educated minority women, additional education is better rewarded in 1990 than in 1972 for all race/sex groups. Interestingly, education was more steeply rewarded among women than men in both years, indicating that in terms of educational differentials, men's wages have shifted toward the steeper profile of women's wages. The impact of these increasing returns to education on overall wage inequality could be mixed, since growth in higher-education differentials for women could reduce earnings inequality because of their generally lower wages.

The other explanatory variable that is estimated quite differently for each race/sex group is returns to potential experience. "Potential" is stressed here because actual experience levels associated with years of potential experience may vary sharply, particularly for women in the early years of the period. One difficulty with experience returns is pinpointing the size and location of the important differences that contribute to earnings inequality. By plotting the full experience–earnings profiles for each of the groups, the scale of the differences can be evaluated at any level of potential experience. Figures 1 and 2 describe the rates of return to experience at the characteristic means of the race/sex groups for 1972 and 1990. Notice that both figures indicate substantial differences; however, the plots converge noticeably over the period.

These two explanatory variables are the most obvious reasons to embark on a decomposition that can account for differing wage structures by race/sex groups, but other differences also exist (for example, demographic differences in the industry-specific terms).

The Decomposition

This paper applies an alternative inequality decomposition that utilizes our understanding of the sources of earnings differences. It uses estimates from standard semilog earnings models to separate earnings into additive components, which can then be evaluated as separate earnings factors. This approach offers several advantages: First, the decomposition can be based on models that have long been accepted by labor economists as reasonably accurate representations of individual earnings. Second, inequality can be speedily decomposed into many categories. Third, inequality can be decomposed according to both discrete and continuous variables.

The decomposition is specified by the following, where Y_i is actually the sum of k component incomes measured in logs, and Y_i^k identifies the k^{th} income component:

$$(6)\quad \sigma^2(Y) = \sum_{i=1}^{N} \frac{(Y_i - \mu)}{n} \sum_{k=1}^{K} Y_i^k$$

$$= \sum_{k=1}^{K} \text{cov}\,(Y_k, Y) \equiv \sum_{k=1}^{K} S_k^*(\sigma^2).$$

The term $S_k^*(\sigma^2)$ follows Shorrocks' (1982) notation for the k^{th} earnings-component-decomposition term of the variance (σ^2), which is measured by the covariance of the income components and total incomes. Shorrocks develops the variance for expository purposes only and does not discuss the variance of log earnings (LV), since it does not satisfy the principle of transfers—a criticism that Creedy (1977) has shown to be irrelevant within the ranges of income or earnings seen in developed economies.[8]

For the simplest case where the earnings factors can be described as partitioned and complete sets of dummy variables, the decomposition on those factors is equivalent to between-group components of a subgroup decomposition on those subgroups. Consider a population that can be divided into N subgroups. Standard coding of the dummy variables results in an X matrix of

$$X = \begin{bmatrix} i_1 & 0 & 0 & \ldots & 0 \\ i_2 & i_2 & 0 & & 0 \\ i_3 & 0 & i_3 & & 0 \\ \vdots & & & & \vdots \\ i_N & 0 & \ldots & & i_N \end{bmatrix},$$

where i_j represents vectors of ones of length n_j, which is the number of members in group j.[9] This matrix excludes the first group from the dummy variables to avoid linear dependence with the intercept. Regression of a vector $Y = (y_i)$ on X results in the following coefficients of β and predictions of $X\beta$:

$$\beta = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_n - \bar{y}_1 \end{bmatrix} \quad \text{and} \quad X\beta = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix},$$

where $\bar{y}_j$ is a vector of the j^{th} group mean. Note that this regression is just another way to calculate the group means.

Treating $X\beta$ and $Y - X\beta$ as factors of the total (Y) and applying the formula for factor decomposition of the variance (equation [6]) result in a standard variance decomposition by subgroups:

$$(7)\quad \text{cov}(X\beta, Y) = \sum_{i=1}^{N} \left[\frac{1}{N}\, \bar{y}_{j(i)}\, y_i - E(\bar{y}_{j(i)})\bar{y} \right]$$

$$= \sum_{j=1}^{J} \left[\frac{1}{N}\, \bar{y}_j \sum_{i \in j} y_i \right] - \bar{y}^2$$

$$= \sum_{j=1}^{J} \frac{n_j}{N} (\bar{y}_j^2 - \bar{y}^2).$$

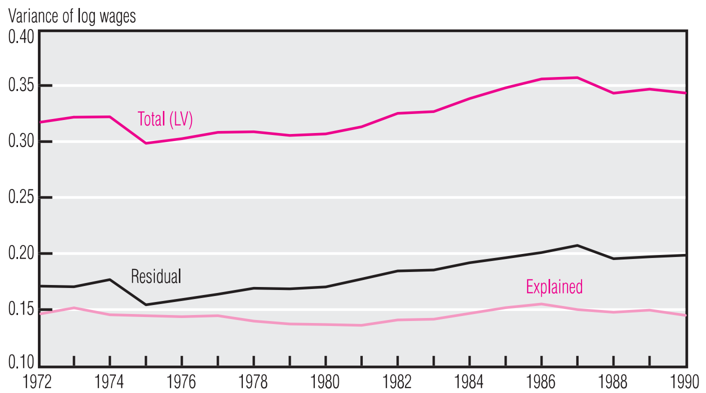
If Y is wages measured in logs, then the final term of equation (7) is the sum of j between-group terms of the subgroup decomposition of the LV. The within-group portion is simply var (Y) – cov$(X\beta, Y)$.

8 The principle of transfers requires the inequality measure to increase whenever income is transferred from a poorer to a richer person.

9 The estimations reported later in this paper are based on regressions in which the dummy variables are effects coded. With effects coding, the coefficients express the difference between group k's mean and all groups' mean wages. The results reported here hold for both standard dummy-variable coding and effects coding, as long as the dummy variables serve as a complete set of group-specific intercepts.

FIGURE 3

Total, Explained, and Residual Earnings Inequality



SOURCE: Author's calculations.

The weighting scheme is identical in the simplest case, but it is easy to move the component decomposition toward simultaneous estimation of a variety of factors affecting earnings. Accounting for covariances can be described as controlling for other effects in a regression framework. The value of this can be seen by considering an example. Wages in many service industries have remained steady or have grown relative to manufacturing wages.[10] This could be due to service industries paying a greater industry differential or to their hiring more-skilled workers (in particular, more-educated workers). In either case, controlling for the level of education and experience in the service workforce would identify lower relative service industry wages. A traditional subgroup decomposition on industries would miss the shift in wage differentials that is hidden by skill upgrading in this sector.

II. Inequality Decompositions

Figure 3 shows the degree to which the model is able to predict observed inequality differences. This is the simplest decomposition possible, but it provides information on how completely the model represents the data. If this were a single-equation model, the percentage of predicted inequality explained would be equal to the R^2 of a regression. Thus, earnings models should not be expected to describe all (or even most) of the variation in earnings when a plethora of important but unobserved individual differences is not taken into account. For 1975, the model predicted an inequality level of 0.1382, which is 46.29 percent of all variation in log earnings. The model predictions of the economywide LV level are relatively stable at around 0.14. However, inequality due to the residual widens throughout the period; thus, the model explains a declining share of the LV of wages.

In addition, the shift in imputation techniques used by the Census Bureau appears to be concentrated in the residuals, which fall from their trend in 1975. At this level of decomposition, the trend in the observed portion of earnings inequality is maintained through the switch in techniques, while the residual portion is dramatically altered.

Factor Shares of Explained Earnings Inequality

This earnings-component-based method of decomposition can be easily applied to any collection of the model's set of variables. Although the overall model's explained inequality changed little from 1972 to 1990, the effects of certain worker characteristics rose or fell rapidly. The results of decomposing the model's estimates into categories are shown in table 4. The experience group includes the quartic terms of potential experience. The education group includes the dummy variables for high school dropout, some college, college graduate, and postgraduate. The race/sex group is implied by the constants of the race/sex-group earnings equations, which are the baseline earnings of individuals of that group after controls have been applied. The industry group includes the 38 industry dummy variables. The region group includes dummy variables for the nine U.S. census regions. The estimated wage effects ($X\beta$) are calculated for each group of variables from annual regressions.

The experience group is a key factor in the explained variation early in the period, reaching 0.0715 LV (or 49.2 percent of explained inequality) in 1974, but its influence declines

10 Average hourly earnings for manufacturing workers fell from \$8.33 in 1970 to \$8.07 in 1990 (1982 dollars). Over the same period, service industry wages rose from \$6.99 to \$7.39 (see 1992 *Statistical Abstract of the United States,* table 650, p. 410).

TABLE 4

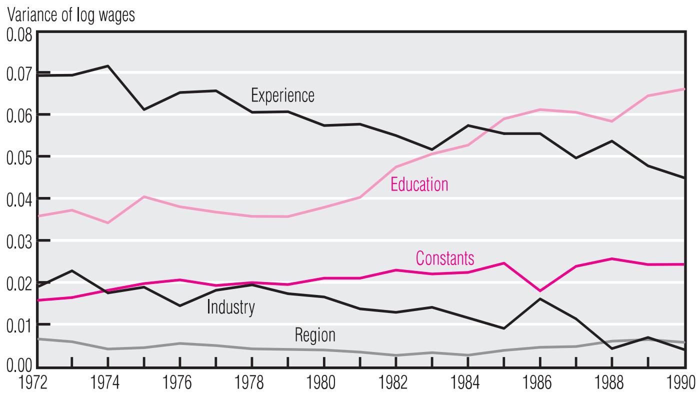
Estimated Earnings Components with Independent Race/Sex Groups

	Levels of Explained Inequality When Factor Returns Differ by Race/Sex Groups					
Year	Total Explained	Experience	Education	Race/ Sex	Industry	Region
1972	0.1463	0.0692	0.0358	0.0191	0.0157	0.0065
1973	0.1514	0.0693	0.0372	0.0227	0.0164	0.0058
1974	0.1454	0.0715	0.0342	0.0175	0.0181	0.0041
1975	0.1444	0.0611	0.0403	0.0188	0.0197	0.0044
1976	0.1436	0.0652	0.0380	0.0144	0.0206	0.0054
1977	0.1446	0.0656	0.0367	0.0181	0.0192	0.0049
1978	0.1397	0.0605	0.0357	0.0194	0.0199	0.0042
1979	0.1371	0.0606	0.0357	0.0173	0.0195	0.0040
1980	0.1366	0.0573	0.0379	0.0165	0.0210	0.0039
1981	0.1359	0.0576	0.0403	0.0137	0.0210	0.0034
1982	0.1408	0.0550	0.0475	0.0129	0.0229	0.0026
1983	0.1415	0.0516	0.0506	0.0140	0.0220	0.0032
1984	0.1465	0.0573	0.0527	0.0115	0.0224	0.0027
1985	0.1517	0.0554	0.0589	0.0090	0.0245	0.0038
1986	0.1550	0.0554	0.0611	0.0160	0.0180	0.0045
1987	0.1499	0.0496	0.0605	0.0113	0.0238	0.0047
1988	0.1477	0.0536	0.0583	0.0042	0.0256	0.0060
1989	0.1496	0.0477	0.0644	0.0068	0.0242	0.0064
1990	0.1449	0.0449	0.0660	0.0040	0.0243	0.0057

SOURCE: Author's calculations.

FIGURE 4

Estimated Earnings Components with Independent Race/Sex Groups



SOURCE: Author's calculations.

thereafter, bottoming out at 0.0449 LV (31.0 percent) in 1990. This is confirmed in figure 4, which compares the relative trends of all five factors. Recalling the experience returns shown in figures 1 and 2, we see that these differences likely reflect the race/sex composition of the labor force as much as they do cohort differences. Previous work with a single-equation model revealed little trend in the experience factor. As the returns to experience converge, but steepen, across demographic groups, the contribution of the earnings factor to inequality declines. This also suggests that it is not experience-profile differences across ages that drive this result, but differences in potential experience returns across groups, because the returns become steeper for all groups by 1990. This is clearly a case where a factor that on its own would contribute to rising inequality (a steeper experience profile) reduces inequality across groups.

Education variables explain a much larger share of the variance of log earnings in recent years. The explained variance accounted for by education dummies rises from a low of 0.0342 LV (23.5 percent of explained inequality) in 1974 to a high of 0.0660 LV (45.6 percent) in 1990. The explanatory power of the education variables increases sharply from the mid-1970s on. Unlike the results for the experience terms, however, rising differentials for all groups (shown in table 3) add to inequality, rather than offsetting other differences. This makes educational levels stand out as a source of inequality that spans different demographic groups in a way that does not ameliorate inequality levels associated with other factors, including the unobserved factors in the residual. It should be noted that while the differentials are certainly important, the fraction of the workforce attaining higher educational levels has also risen.

The race/sex term is defined by the constants of the regression equations. It thus represents baseline differences not associated with return differences on the included factors, rather than an inclusive measure of group differences. It starts with a peak explanatory power of 0.0227 (15.0 percent of explained earnings inequality) in 1973, but by 1990 accounts for only 0.0040 LV (2.7 percent). This dramatic decline, which is spread over the period, has not been noted in previous studies because most researchers either have considered only men or have treated men and women as if they participated in different labor markets. Combined with the reduced effect of experience as a factor in inequality, factors closely related to differences among demographic groups have declined considerably as sources of inequality.

The share of industry variables in explained inequality is not as large or as steeply trended as either the education or race/sex shares. However, the effect of industry wage differentials rises from 0.0157 LV (10.7 percent of explained earnings inequality) in 1972 to 0.0256 LV (17.3 percent) in 1988. The share of inequality represented by the industry factor does little to bolster theories positing that the increase in overall inequality results mainly from industrial shifts. However, unlike in previous studies, the trend in the industry component is noticeably upward and economically significant. Regional differences play a consistently small role in earnings inequality, reaching a low of 0.0026 LV (1.9 percent) in 1982.

Simple calculations from table 4 indicate that trends in some of these factors are quite large. Inequality due to educational differences rose 116.2 percent more than overall earnings inequality among full-time/full-year labor force participants from 1972 to 1990. This implies that if all factors other than education (including the residuals) were held constant over the period, earnings inequality would have risen 16.2 percent more than it actually did. Even the relatively small industry factor grew 33.1 percent as much as overall inequality.

These increases are more than offset by the drop in the experience and race/sex factors. Experience-related inequality declined at a rate equal to 93.5 percent of the increase in overall earnings inequality over the 1972–90 period, and the race/sex factor fell 58.1 percent versus the same measure. These factors, combined with the small regional factor, yield explained inequality levels that actually decrease as overall inequality and inequality associated with education and industry affiliations rise.

Fixed-Return Comparisons

A valuable extension of the preceding analysis is to separate the effects of population shifts from the effects of changes in returns to worker characteristics. Basic shift/share analysis, in which a population having given characteristics is adjusted in order to isolate the population effects, cannot be applied to this decomposition because the correlations of individual characteristics at the observation level are critical. Shift/share analysis implicitly assumes that the nature of the correlations stays constant.

A related approach is to contrast the explained inequality level under the restriction that the estimated coefficients are constant in all years. Here, the restricted case is referred to as fixed-return estimates because the coefficients represent the amount a hypothetical average individual is paid for having that characteristic. This comparison can isolate the effects of changes in rates of return paid to earnings factors from the changing distribution of those factors. Much as in shift/share analysis, in addition to the returns and quantity terms, there is a covariance between the two terms that is assumed to be zero. This allows for the simple separation

$$S_k^*(\sigma^2) = \operatorname{cov}(X_k \beta_k, Y_i) = \operatorname{cov}(X_k \tilde{\beta}_k, Y_i) + \operatorname{cov}[X_k(\beta_k - \tilde{\beta}_k), Y_i], \quad (8)$$

where $\tilde{\beta}_k$ represents any desired value of the coefficient vector for the k^{th} factor.

Table 5 shows the difference between the restricted (coefficients maintained at 1972 levels) and unrestricted inequality components over time. The difference between the two estimates equals the final term in equation (8), which is an inequality-weighted measure of the difference between coefficients. A positive value indicates that allowing the coefficients to vary increases inequality; a negative number implies reduced disparity when coefficients are allowed to change.

If all returns to worker characteristics were held at their 1972 levels, the explained level of earnings inequality would have been slightly lower in 1990 than in 1973. This suggests that overall shifts in the composition, without any change in the earnings functions, has raised earnings inequality. Referring again to table 5, with constrained rates of return, earnings inequality would have been higher, with 1972's return levels, from 1974 to 1983. The reversal of this result is due to sharply rising returns to education in the 1980s. The largest differences, and therefore the largest return-related shifts, occurred in the education-, race/sex-, and industry-related components.

The change in experience-related inequality is 73 percent larger when determined without any change in relative returns, whereas race/sex-related inequality drops off only 49 percent as much. Education-related inequality is even more affected by shifts in the number of workers at various schooling levels than by shifts in the returns for increases in that component (64 percent of the change would have occurred with no change in returns). By contrast,

TABLE 5

Effect of Holding Returns Constant with Independent Race/Sex Groups

	Increase in Factor Estimates with Flexible Returns					
Year	Total Explained	Experience	Education	Race/ Sex	Industry	Region
1972	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1973	0.0029	–0.0025	0.0007	0.0033	0.0018	–0.0005
1975	–0.0003	0.0004	–0.0014	–0.0013	0.0034	–0.0014
1976	–0.0009	–0.0064	0.0023	0.0009	0.0036	–0.0013
1977	–0.0009	–0.0029	0.0012	–0.0031	0.0041	–0.0002
1978	–0.0008	–0.0032	–0.0001	0.0005	0.0023	–0.0002
1979	–0.0024	–0.0065	–0.0007	0.0019	0.0034	–0.0005
1980	–0.0029	–0.0050	–0.0010	0.0000	0.0036	–0.0005
1981	–0.0035	–0.0073	–0.0001	–0.0002	0.0047	–0.0008
1982	–0.0041	–0.0060	0.0008	–0.0027	0.0047	–0.0010
1983	–0.0008	–0.0069	0.0041	–0.0029	0.0064	–0.0015
1984	0.0002	–0.0086	0.0055	–0.0010	0.0056	–0.0013
1985	0.0027	–0.0038	0.0056	–0.0034	0.0061	–0.0018
1986	0.0062	–0.0046	0.0099	–0.0055	0.0076	–0.0011
1987	0.0070	–0.0044	0.0099	0.0019	0.0005	–0.0009
1988	0.0044	–0.0080	0.0088	–0.0028	0.0069	–0.0006
1989	0.0034	–0.0022	0.0071	–0.0091	0.0075	0.0001
1990	0.0050	–0.0063	0.0100	–0.0060	0.0067	0.0007

SOURCE: Author's calculations.

virtually all of the rise in the industry component is driven by shifts in industry wage differentials, rather than by shifts in industry employment shares.

The fact that much of the change in earnings inequality occurs without changes in relative earnings is significant. A large part of the change in demographics is predictable. We know the basic characteristics of people poised to enter the labor force, and we can guess that trends in industry employment shares are likely to continue for several years. The retiring labor force in the United States is more male, more white, less educated, and more likely to work in manufacturing industries. Replacing these workers implies a continuation of offsetting compositional changes on the earnings inequality of the workforce that, depending on which effect dominates, will determine much of the inequality trend into the next decade.

III. Conclusion

Stepping back from the technical details of the decompositions, we can see the complexity involved in addressing total earnings inequality as a public policy issue. While policies might be easily structured to benefit particular groups of workers, important covariances in earnings factors across groups can lead to changes in overall inequality that are either positive or negative. The decomposition employed in this paper can be used to verify or alter findings based on studies of single demographic groups.

Notably, the growing importance of educational differences is verified across all four demographic groups examined here, despite their widely varying schooling levels. The rise in education-related inequality, which is generally ascribed to rising returns, appears to be more than 50 percent determined by the size of the highly educated labor force, at least in this sample. Neglecting the participation of a growing fraction of the labor force may have caused previous researchers to focus excessively on shifts in returns.

The analysis also establishes the direct role of changing workforce demographics. Race/sex differentials have contributed far less to recent inequality levels than was historically the case, masking part of the widening disparity in other factors. These trends are driven both by changes in relative pay rates and shifts in the composition of the labor force. The largest factor is differing rates of return on potential experience by race/sex group. Declines in this factor have resulted primarily from changing participation rates, not from shifts in the observed experience–earnings profiles.

The decompositions presented here generally point to a larger role for the composition of the full-time/full-year workforce than has previous research. Policy prescriptions based on the existing literature tend to ignore the effects of this striking change. While such remedies may still be appropriate, the fact that much of the inequality trend has been driven by changes in the composition of the U.S. labor force suggests that policies which alter the returns to schooling or other human capital factors will take a long time to work. One reason is that the composition changes realized over the last decade are likely to continue, if only because entering generations are replacing retiring workers who possess characteristics much more typical of the earliest periods of this analysis.

Appendix: The Data Set

The data set is derived from the March Current Population Surveys (CPS) spanning the years 1973 to 1991. Every month, the U.S. Census Bureau interviews about 58,000 households (including approximately 122,000 persons age 14 and over) as part of the CPS. Each sample is designed to be representative of the civilian, noninstitutional population. The March surveys throughout this period include information on individuals' personal characteristics (age, sex, race, and education) and on their residence and employment during the previous year (total wages and salaries, weeks worked, hours worked per week, industry, and occupation). These features have made the March supplement the primary data source used in earnings distribution analyses.

I selected individuals who showed strong attachment to the labor force. The sample includes civilians over age 16 who are not self-employed and who missed no weeks of work because of schooling or retirement. It is further limited to workers who reported being in the labor force (working or unemployed) at least 39 weeks and who worked full time (at least 35 hours per week) in the previous year. Although designed to be similar to the sample used by Juhn, Murphy, and Pierce (1993), mine includes both male and female workers of all races in order to paint a more complete picture of the labor market.

Certain adjustments to the earnings data were also necessary. Top-coded data were assigned the truncated mean earnings implied by a Pareto distribution based on the highest reported earnings. Observations with real weekly wages of less than half the 1982 minimum wage for a full-time job were dropped because they are likely to be faulty. Juhn, Murphy, and Pierce show that differences in the imputation techniques used by the Census Bureau can alter wage inequality, but that these differences are largely limited to extremes of the distribution. The biggest switch occurs between 1974 and 1975 and is visible in the decompositions reported here. To isolate the conclusions of this paper from the issues that affect the fringes of the distribution, the analysis was also completed with a sample from which the top and bottom 5 percent of earners were removed.[11] There were no differences in the truncated sample analysis that would alter the conclusions of this paper.

The Census Bureau changed its industry codes twice during the sample period. However, the basic structure of the industry coding system was not altered at the two-digit level and could be mapped into consistent two-digit Standard Industry Codes. I aggregated some of these codes in order to reduce the number of industries to a manageable number (39) and to increase the cell sizes for small industries.

11 Neither top-coded data nor subminimum wage earnings were ever more than 5 percent of my sample.

References

Blackburn, M.L. "What Can Explain the Increase in Earnings Inequality among Males?" *Industrial Relations*, vol. 29, no. 3 (Fall 1990), pp. 441–56.

Burtless, G. "Earnings Inequality over the Business and Demographic Cycles," in G. Burtless, ed., *A Future of Lousy Jobs?* Washington, D.C.: The Brookings Institution, 1990, pp. 77–117.

Creedy, J. "The Principle of Transfers and the Variance of Logarithms," *Oxford Bulletin of Economics and Statistics*, vol. 39, no. 2 (May 1977), pp. 153–58.

Eberts, R.W. "Accounting for the Recent Divergence in Regional Wage Differentials," Federal Reserve Bank of Cleveland, *Economic Review*, vol. 25, no. 3 (Quarter 3 1989), pp. 14–26.

Freeman, R.B. "How Much Has De-Unionization Contributed to the Rise in Male Earnings Inequality?" National Bureau of Economic Research, Working Paper No. 3826, August 1991.

________. "Toward an Apartheid Economy?" *Harvard Business Review*, September/October 1996, pp. 114–21.

Heckman, J., and J. Scheinkman. "The Importance of Bundling in a Gorman–Lancaster Model of Earnings," *Review of Economic Studies*, vol. 54, no. 2 (April 1987), pp. 243–55.

Juhn, C., and K.M. Murphy. "Wage Inequality and Family Labor Supply," *Journal of Labor Economics*, vol. 15, no. 1, part 1 (January 1997), pp. 72–97.

__________, __________, and B. Pierce. "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, vol. 101, no. 3 (June 1993), pp. 410–42.

Karoly, L.A. "Changes in the Distribution of Individual Earnings in the United States: 1967–1986," *Review of Economics and Statistics*, vol. 74, no. 1 (February 1992), pp. 107–15.

Levy, F., and R.J. Murnane. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, vol. 30, no. 3 (September 1992), pp. 1333–81.

Mincer, J. "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, vol. 66, no. 4 (August 1958), pp. 281–302.

________. *Schooling, Experience, and Earnings.* New York: National Bureau of Economic Research, 1974.

Murphy, K.M., and F. Welch. "Empirical Age–Earnings Profiles," *Journal of Labor Economics*, vol. 8, no. 2 (April 1990), pp. 202–29.

________, and ________. "The Structure of Wages," *Quarterly Journal of Economics*, vol. 107, no. 1 (February 1992), pp. 285–326.

Oi, W.Y. "Labor as a Quasi-Fixed Factor," *Journal of Political Economy*, vol. 70, no. 6 (1962), pp. 538–55.

Shorrocks, A.F. "Inequality Decomposition by Factor Components," *Econometrica*, vol. 50, no. 1 (January 1982), pp. 193–211.

_______. "Inequality Decomposition by Population Subgroups," *Econometrica*, vol. 52, no. 6 (November 1984), pp. 1369–85.

Smith, J., and F. Welch. "Inequality: Race Differences in the Distribution of Earnings," *International Economic Review*, vol. 2, no. 2 (June 1979), pp. 515–26.

The Risk Effects of Bank Acquisitions

by Ben Craig and João Cabral dos Santos

Ben Craig and João Cabral dos Santos are economists at the Federal Reserve Bank of Cleveland. The authors appreciate the helpful comments of William Osterberg and an anonymous referee. They also thank Jennifer Carr, Jean McIntire, and especially Rebecca Wetmore Humes and Sandy Sterk for their dedicated research assistance.

Introduction

Since the early 1980s, the U.S. banking industry has seen a strong trend toward consolidation, partly because of state regulatory changes permitting out-of-state bank acquisitions. There were 6,157 bank mergers and acquisitions (M&As) between 1981 and 1994 (Rhoades [1996]). Consolidation of this magnitude has brought significant changes to the banking sector that are in themselves worth investigating. By identifying these changes, we also gain valuable information about the ongoing wave of M&As that began with enactment of the Interstate Banking and Branching Efficiency Act in 1994.[1] Moreover, M&As require regulators' preapproval, and information on the likely effects of such changes can be useful in the approval process.

A prime objective of research on bank M&As has been to identify motives for consolidation. Such motives as scale economies, scope economies, and managerial X-efficiencies have been studied extensively.[2] However, less attention has been given to the other two most frequently suggested motives for bank M&As: risk diversification, and the wish to become "too big (or too important) to fail."[3] The present paper contributes to the literature by evaluating the importance of the risk diversification motive. Our study considers only bank acquisitions, which differ from mergers in that the acquired bank continues to operate as an institution after being acquired; it does not lose its charter. This focus on acquisitions allows us to identify how each party—the acquirer and the acquired bank—affects the risk of failure of the newly formed banking organization.

The paper proceeds as follows: The next section discusses the importance of the risk diversification motive and defines our contribution to the related literature. Section II presents

■ 1 The Act defined nationwide standards for a bank holding company (BHC) to acquire a bank in any state. Moreover, beginning June 1, 1997, BHCs were allowed to convert their bank subsidiaries into a network of branches, provided that these banks' home states had not opted out of the Act's branching provision.

■ 2 Useful reviews of the literature on economies of scale and scope are presented by Clark (1988) and Mudur (1992). Berger, Hunter, and Timme (1993) review the literature on X-efficiencies.

■ 3 Hunter and Wall (1989) and Boyd and Graham (1991), among others, raise the possibility that banks seek to become larger in order to increase their deposit insurance subsidy by being considered too big (or too important) to fail.

the measures of risk and our method for identifying the acquisition effects. Section III describes our sample of bank acquisitions, and section IV presents the results. The paper closes with some final remarks on the policy implications of our study.

I. The Risk Diversification Motive

In the debate on the risk diversification motive for bank M&As, some argue that banks choose targets that allow for a significant reduction in their risk exposure. Others suggest that, because of the moral hazard created by deposit insurance, a merger or acquisition gives the acquiring bank a good opportunity to increase its deposit insurance subsidy either by increasing its risk exposure or by attempting to become too big to fail. Still others say that risk considerations play no significant role in banks' merger policies.

Despite the importance of this debate, little is known about the risk effects of bank M&As. On the one hand, the research on postmerger effects has concentrated on performance (profits and costs) and on the changes in asset management (composition of the bank's portfolio of assets) resulting from the merger or acquisition. On the other hand, research on the risk effects of acquisitions has focused on combinations of banks and nonbank financial firms.[4]

The indirect evidence on the importance of the risk diversification motive is somewhat mixed. For example, Lawrence (1967), Talley (1971), Ware (1973), and Hobson, Maston, and Severiens (1978) find that acquired banks tend to adjust the composition of their portfolios by switching out of U.S. government securities in to loans and state and local government securities. These studies report mixed effects on the acquired bank's capital–asset ratio. Rhoades (1987), Fraser and Kolari (1987), and Beatty, Santomero, and Smirlock (1987) find a negative relationship between the merger premium and the target bank's capital–asset ratio. Craig and Santos (1996) show that regardless of the acquired bank's characteristics, the acquiring institution changes the target bank's asset composition so that the resulting organization becomes a bigger version of the acquirer. When Rose (1989) asked managers of banks involved in mergers to indicate the motives for consolidation, risk reduction was one of the least frequently mentioned responses.

The only study we know that directly compares the importance of the risk reduction motive with the deposit subsidy enhancement motive is Benston, Hunter, and Wall (1995). They conjecture that acquirers seeking to reduce risk should be willing to pay a premium for target banks that will lower the risk of the new banking organization. Under this hypothesis, there should be a negative relationship between the purchase premium and the target's expected contribution to the risk of the new organization, which they proxy by the variance of the target's return on assets and the covariance between the acquirer's and the target's return on assets, both computed prior to the acquisition. If the acquirer uses the acquisition to increase its deposit insurance subsidy instead, it can accomplish this objective either by becoming too big to fail or by increasing its risk exposure. Under this hypothesis, the purchase premium should be positively related to the two measures of risk already mentioned and to the acquirer's risk (as measured by the variance of its return on assets), and negatively related to the ratio of the acquirer's book value of equity to its total asset value. Benston, Hunter, and Wall contend that their results are consistent with the hypothesis of reducing risk and inconsistent with that of enhancing the deposit insurance subsidy.

The present study follows a different route for evaluating the importance of the risk diversification motive for bank acquisitions. We conjecture that if it were an important motive, a reduction in risk should follow the acquisition. That is, the postacquisition risk of the newly formed organization should be lower than the preacquisition risk of the acquiring bank holding company (BHC).

In assessing the importance of risk diversification, we use several measures to compare the postacquisition risk of the newly formed banking organization with the preacquisition risk of the acquiring BHC. To determine the source of risk effects resulting from the acquisition, we also compare the same measures of risk before and after acquisition for both institutions. Furthermore, we compare the new banking organization's risk with that of the hypothetical banking organization that would result from the preacquisition aggregation of the acquiring and the acquired banks. The purpose of this comparison is to gain information on how consolidation has affected the banking industry's overall risk.

4 See, for example, Litan (1987), Santomero and Chung (1992), and Boyd, Graham, and Hewitt (1993).

We complement the above analysis by examining the dynamics of the risk effects caused by bank acquisitions, using a constant sample. That is, for a given time frame, defined around the acquisition date, we consider only acquisitions for which we have observations throughout the entire period, and then study the dynamics of the risk effects within that interval. Thus, we avoid two problems that are frequently encountered in the literature on post-acquisition effects: having a sample whose composition changes over time, and having a sample of acquisitions that all occur at the same time.

II. Method and the Measurement of Risk

We identify the risk effects of bank acquisitions through a two-step procedure. First, we compute the risk for the banking organizations in our sample, both before and after the acquisition. We then normalize this measure by subtracting the mean of the same measure calculated for the set of all banks in the industry, excluding those in our sample. By doing so, we eliminate a time effect—that is, a shock to the risk that prevails in the entire industry during a given period. In the second step, we evaluate the acquisition effect by computing the difference (postacquisition minus preacquisition) between the two industry-normalized measures calculated in the first step. This procedure removes any individual effect, that is, any idiosyncratic risk associated with the banking organizations involved in the acquisition. We then test statistically for whether this difference is zero. A number significantly different from zero indicates that the acquisition caused a change in risk.

In this study, we consider three indicators that are generally adopted in the literature to measure a banking organization's risk. The first two are the standard deviation and the coefficient of variation of a bank's profitability. These measures are computed for both the return on assets (the ratio of net income to total assets) and the return on equity (the ratio of net income to equity capital). The third indicator is what has become known in the literature as the *Z*-score, a measure of a bank's probability of bankruptcy.[5] In addition, because of their importance in defining some of these risk indicators, we study an acquisition's impact on both the profitability (return on assets and return on equity) of the banking organizations involved and the covariance between the profitability of the acquiring and target banks.

The *Z*-score can be defined as follows: Let bankruptcy be the situation in which the bank's equity capital, E, is smaller than its losses, $-\pi$ (since π represents the bank's profits); that is, $E < -\pi$. In addition, let A be the bank's total assets, r the bank's return on assets, $r = \pi/A$, and k the negative of the equity-to-assets ratio, $k = -E/A$. Using these definitions, the bank's probability of bankruptcy can be written as

$$(1) \qquad p(\tilde{\pi} < -E) = p(\tilde{r} < k) = \int_{-\infty}^{k} \phi(\tilde{r})\, d\tilde{r},$$

where $p(\cdot)$ is a probability, $\tilde{\pi}$ and $\tilde{r}$ represent random variables, and $\phi(r)$ is the density function.

If $\tilde{r}$ is assumed to have a normal distribution, then the bank's probability of bankruptcy can be rewritten in terms of the standard normal density, $\Psi(\cdot)$, as

$$(2) \qquad p(\tilde{r} < k) = p(\tilde{r} < k) = \int_{-\infty}^{z} \Psi(\zeta)\, d\zeta,$$

where $\zeta = \dfrac{\tilde{r} - \rho}{\sigma}$ and $z = \dfrac{k - \rho}{\sigma}$, with ρ being the true mean and σ the standard deviation of the random variable $\tilde{r}$.[6]

The *Z*-score, or sample estimate of $-z$ (because $z < 0$), is computed using the sample estimates of ρ and σ. As a result, based on quarterly accounting data, the *Z*-score is defined for each bank as

$$(3) \qquad Z = \frac{\sum_{i=1}^{n} 2n \dfrac{\pi_i}{A_i + A_{i-1}} + \sum_{i=1}^{n} n \dfrac{E_i + E_{i-1}}{A_i + A_{i-1}}}{\left[\sum_{i=1}^{n} \left(2 \dfrac{\pi_i}{A_i + A_{i-1}} - \sum_{i=1}^{n} 2n \dfrac{\pi_i}{A_i + A_{i-1}} \right)^2 \dfrac{1}{n-1} \right]^{\frac{1}{2}}},$$

where the stock variables, equity, and assets are measured at the midpoint of the period, and n is the number of periods considered in the sample.[7]

5 For detailed analyses of this measure of risk, see Meinster and Johnson (1979) and Boyd, Graham, and Hewitt (1993).

6 Because of Chebyshev's inequality, we know that regardless of the distribution of $\tilde{r}$, as long as ρ and σ exist, the upper bound to the bank's probability of bankruptcy is

$$p(\tilde{r} \leq k) \leq \left(\frac{\sigma}{\rho - k} \right)^2 = \frac{1}{z^2}.$$

7 Because we consider only the acquisition of banks, and not BHCs, the computation of the *Z*-score for the target bank is straightforward. The same holds when the acquisition is made by a BHC that owns only one bank. When a multibank BHC makes the acquisition, the *Z*-score is computed for the hypothetical bank created as the sum of the banks in the acquiring BHC.

TABLE 1

Sample Composition

		Number of Banks in the BHCs after the Latest Acquisition									
		2	3	4	5	6	7	8	9	10	11
Number of Banks in the BHCs before the First Acquisition	1	196	7	—	—	—	—	—	—	—	—
	2	—	24	1	—	—	—	—	—	—	—
	3	—	—	7	1	—	—	—	—	—	—
	4	—	—	—	3	1	—	—	—	—	—
	5	—	—	—	—	2	—	—	—	—	—
	6	—	—	—	—	—	1	—	—	—	—
	7	—	—	—	—	—	—	—	1	—	—
	…	—	—	—	—	—	—	—	—	—	—
	10	—	—	—	—	—	—	—	—	—	1

SOURCE: Authors' calculations.

The *Z*-score is an indicator of a bank's probability of bankruptcy in the sense that it estimates the number of standard deviations below the mean that the institution's profits would have to fall before its equity became negative. Thus, the smaller the value of *Z*, the larger the bank's risk of failure. Looking at the definition of *Z*, we observe that its value depends positively on the bank's profitability (measured by its return on assets) and capital–asset ratio, and negatively on the variability of the bank's profits (measured by the standard deviation of its return on assets).

In the second step of our procedure, we identify the acquisition effects by comparing before and after measures of risk and profitability for different banking organizations. By subtracting the preacquisition measure from the postacquisition measure, we can gauge the consequences of acquisition for the individual bank (or group of banks) affected. The individual differences are averaged and a standard t-test is used to check whether the means equals zero.

We account for the market effects on our risk measures by normalizing them with corresponding statistics for the banking industry as a whole. Each risk measure is computed as a deviation from the industry average for the same time period. Take, for example, the case of the *Z*-statistic in the 4 by 16 sample. As we did with the banks in our sample, we computed the *Z*-statistic for each bank outside our sample using quarterly data for the year before the acquisition quarter and for each of the four years after the acquisition. The average for the industry (which excludes the banks in our sample) is subtracted from the corresponding statistic for the acquisition pair. This removes effects that influence not only the acquisition pair, but also the industry as a whole. Thus, each measure of risk is expressed as a deviation from the industry average.

Because the sample sizes were large—173 or 201 acquisition pairs, depending on the time frame—standard central-limit-theorem results could be expected to hold fairly closely. It is important to note that the measures of risk all use sample means that are calculated separately for each period and each bank. Furthermore, the sample size is based on the number of individual pairs, not on the number of pair–quarter observations. While this procedure is robust to changing sample means or an unspecified stochastic process in the return (such as a first-order autoregressive process), our test is conservative in the sense that it tends to reject less often than a more fully specified statistical model. However, because our results generally rejected at a high level or did not reject at any reasonable level of significance, we report tests for the robust statistical model.

III. Sample Construction

Our study reports results only for acquisitions in which a bank continues to exist after being acquired by another banking organization. Furthermore, we restrict our sample to acquisitions made by one-tier BHCs, which own banks but do not own other BHCs. The data for this study were obtained from banks' quarterly Reports of Condition and Income (call reports) for the first quarter of 1984 through the last quarter of 1993.

Table 1 summarizes our sample of 256 bank acquisitions. Note that the largest number of acquisitions (196) was made by BHCs that had one bank and acquired a second during the sample period. Note also that several BHCs made more than one acquisition during our sample period. For example, there were seven BHCs that had one bank when the period began and later acquired a second bank and then a third.

Table 2 contains some descriptive statistics on the ratio of acquired banks' assets to those of the largest bank in the acquiring BHC. This information is arranged according to the number of banks in the acquiring BHC. For example, in nine acquisitions, BHCs with three banks acquired a fourth and, on average, the acquired banks' assets were 46 percent of those of the largest of the three banks in the acquiring BHC.

TABLE 2

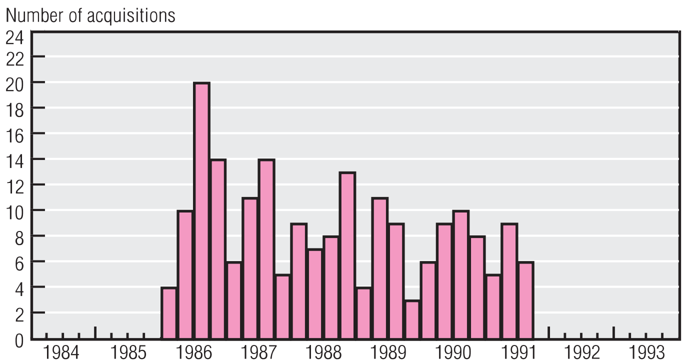
Acquired Bank's Assets/Assets of Largest Bank in Acquiring BHC

	Number of Banks in the Acquiring BHC									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
N[a]	203	32	9	5	3	1	1	1	—	1
Mean	0.56	0.63	0.46	0.18	0.31	0.49	0.11	0.06	—	0.06
Median	0.43	0.43	0.30	0.17	0.31	0.49	0.11	0.06	—	0.06
CV [b]	0.53	0.72	0.36	0.11	0.19	—	—	—	—	—

a. Number of acquiring BHCs.
b. Coefficient of variation.
SOURCE: Authors' calculations.

FIGURE 1

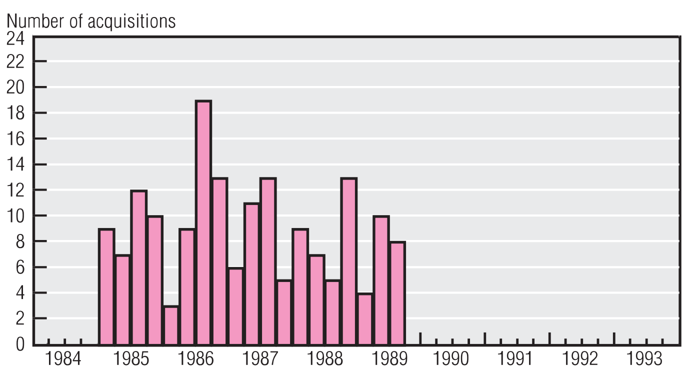
Distribution of Bank Acquisitions in the 8 by 8 Time Frame



SOURCE: Authors' calculations.

FIGURE 2

Distribution of Bank Acquisitions in the 4 by 16 Time Frame



SOURCE: Authors' calculations.

As expected, acquired banks are substantially smaller on average than their acquirers.

We conduct our tests on samples of bank acquisitions that are held constant within a given time frame, that is, an interval with a set number of quarters before and after the acquisition date. By constant sample, we mean a sample of acquisitions, each of which satisfies two criteria: first, availability of data on both parties involved in the acquisition throughout the entire time frame; second, structural constancy, meaning that the parties involved made no further bank acquisitions or sales during the period defined by the time frame.

This study employs two time frames. The first includes acquisitions for which we have the appropriate data for eight quarters before and eight quarters after the acquisition. In this case, we compare the risk measures for these two intervals. The second time frame includes four quarters prior to the acquisition and 16 quarters after it. Here, we compare the measures of risk for the four quarters prior to the acquisition with the same measures computed for the four quarters of the first year after acquisition, for the four quarters of the second year, and so on, through the fourth year after acquisition. This broader horizon allows us to identify the dynamics of the risk effects during the four years following the acquisition.

Given these conditions, we were able to consider 201 bank acquisitions in the 8 by 8 time frame and 173 in the 4 by 16 time frame. The temporal distribution of these acquisitions is presented in figures 1 and 2, respectively.

IV. Results

Before presenting the results of our tests on the risk effects of bank acquisitions, it is important to compare our sample of banks to the industry as a whole, before and after acquisition. The acquirers in our sample have a higher return on assets and a higher return on equity than the industry average in both periods. Acquired banks, however, according to both measures of profitability, are below the industry average before being acquired, but above it afterward. Furthermore, the improvement in their profitability increases with time.

With respect to risk, our sample of acquirers is made up of banks that appear safer than the industry average, both before and after making their acquisitions. This is clearest when risk is measured by the standard deviations of the return on assets and on equity. It is less clear when risk is measured by the *Z*-score, because in some periods the acquirers do not differ statistically from the industry. The results are inconclusive when risk is measured by the coefficients of variation of the returns on both assets and equity. Target banks are riskier than the industry before being acquired, according to both the standard deviation measures and the Z-score. Afterward, their difference from the industry as a whole shrinks; in fact, they become relatively safer, according to the standard deviation measures of risk. As with the acquirers, no clear results are obtained from the targets' coefficients of variation.

The results of our tests for the effects of bank acquisitions are shown in the appendix (tables A1 to A4). In each time frame, every cell is associated with every measure of risk or profitability and each type of institution. The cell's first line represents the mean of the "after" minus the "before" difference for the statistic associated with that column. The former is defined as the postacquisition difference for that statistic between the sample and the control group, that is, the industry as a whole (excluding the banks in our sample). The "before" difference is defined the same way, but it is computed before the acquisition. The second line represents the p-value for the null hypothesis that the "after" minus the "before" difference is zero. The p-value is calculated using the standard t-distribution. For example, the cell in table A1 associated with acquired banks and the standard deviation of the return on assets, S_A, indicates that the mean of this measure of risk, after adjusting for the industry effect, was smaller (by an amount equal to 0.0011) for the period encompassing eight postacquisition quarters than the mean that existed for the same length of time before the acquisition. The hypothesis that this number is zero has a p-value of 7.94×10^{-9}, which means that no difference between the two measures would be rejected at any reasonable level of significance. It is evident that the standard deviation of the return on assets for the acquired banks decreases as a result of acquisition.

A detailed analysis of tables A1 through A4 reveals some clear results. First, the new banking organizations that emerge from acquisitions show improved profitability, which increases with time. Indeed, the profitability of these organizations ultimately exceeds that of the associated acquirers prior to acquisition. It also surpasses the preacquisition profitability of the hypothetical institutions that result from the aggregation of the acquirers and the target banks. The improvement is explained by the increased profitability of both the acquirers and the acquired institutions, but particularly by the latter. These patterns are observed for the returns on assets and equity.

Second, the variability of new banking organizations' returns is reduced (as measured by both the standard deviation of the return on assets and the standard deviation of the return on equity) as a result of acquisitions when compared to the corresponding preacquisition variability of both the acquirers' and the hypothetical institutions' returns. The variability of returns decreases over time and is observed for both the acquirers and the acquired. As in the case of profitability, the reduction is more pronounced for the latter.

Third, based on the Z-score, we observe that the new organizations which emerge from acquisitions have a lower probability of failure than the hypothetical preacquisition organizations. Furthermore, this improvement increases over time (recall that the larger the value of the Z-score, the smaller the bank's risk of failure). Acquired banks follow nearly the same pattern. However, acquisitions do not affect the Z-scores of acquirers in a statistically significant way. An identical conclusion is reached when the postacquisition Z-scores of new banking organizations are compared to the corresponding acquirers' preacquisition Z-scores.

Following the leads of other researchers, we have also examined changes in the coefficient of variation and in the covariance. While these results seem to contradict those of other measures, careful consideration of the statistical properties of the two measures leads us to urge caution in using them to evaluate the risk effects of acquisitions. In the case of the coefficient of variation, the lack of a clear result is seemingly related to the fact that random draws of a variable $1/t$, where t is a random variable from a student t-distribution, has a mean that quickly approaches a limiting Cauchy distribution. Our results were strongly influenced in ways that are typical of those caused by the Cauchy distribution's lack of statistical moments. Often, for example, computations of industry means for the coefficient of variation depended completely on the behavior of one or two huge outliers. Even trimming the industry means did not completely solve these problems. Our experience with the coefficient of variation provides a strong cautionary moral for researchers who forget that central-limit theorems depend on the existence of finite moments greater than two. The coefficient of variation behaves too much like a Cauchy variable to depend on such statistical results. For handling this problematic variable, we advise using procedures that take the median as a basis for location estimation.[8]

Similar results were also obtained for the covariance risk measure, although it is unlikely that these are due to observations being drawn from a distribution with no finite moments. In this case, we did not normalize for industry averages. The industry as a whole consists of individual banks rather than pairs, so that an industry mean could not be subtracted. Given that the effects of acquisition on the covariance measure are slight, our results seem to be influenced most by random noise and therefore provide little in the way of information about small effects that may be due to acquisitions. Clearly, acquisition does not change the covariance measure enough to make it detectable using our methods.

In sum, we find that acquisitions have a positive impact on the profitability of participating institutions—particularly acquired banks—and that it increases over time. This improvement, in turn, reduces the variability of the institutions' profits in a similar fashion. These results, a priori, would appear to contribute to participating banks' reduced probability of failure, as measured by the *Z*-score. In fact, this is true for acquired banks and for the new banking organizations that result from acquisitions when these are compared to hypothetical preacquisition organizations. However, this reduction is not observed when the comparison is made with preacquisition acquirers' risk, nor is it seen for acquirers. These results may be explained by the difference in capitalization between acquirers and acquired banks. They seem compatible with a situation in which the capitalization of acquired banks is smaller than that of their acquirers prior to the acquisition, but is improved as a result of the acquisition.

What do our findings tell us about the risk diversification motive for bank acquisitions? They appear to disprove the notion that banks use acquisitions as a means of boosting their deposit insurance subsidy by increasing their risk exposure. In this respect, our results accord well with those of Benston, Hunter, and Wall (1995). However, our evidence is mute about the other frequently suggested route for increasing the deposit insurance subsidy—becoming too big to fail.

Although they indicate that acquisitions have increased the banking industry's profitability and reduced its risk, our findings do not seem to show that risk diversification is in itself a major force driving bank acquisitions. If that were the case, we would have observed a distinctly reduced risk for the new banking organizations that emerge with acquisitions. Our sample of bank acquisitions shows a clear reduction for some measures of risk, but no clear effect for others. Furthermore, when we compare the postacquisition risk measures of the newly formed banking organizations with the preacquisition risk of the corresponding hypothetical banking organizations, we find a more pronounced reduction than when we compare those same measures with the preacquisition risk of the associated acquirers.

V. Final Remarks

Recent consolidation in the banking industry is producing less risky organizations. At the same time, it is creating larger institutions for which the too-big-to-fail policy is potentially more valuable. The moral hazard of that policy creates a widely recognized distortionary subsidy. In the past, regulations limiting bank acquisitions, especially interstate M&As, made it difficult to exploit this subsidy. Recent deregulation allowing the development of nationwide banking has made it easier for banks to diversify their risks, but it has also made it easier for them to grow. In other words, the perception that banks could become too big (or too important) to fail is more pertinent than ever before. As a result, it is imperative that banking supervisory agencies avoid sending signals that might reinforce this perception.

8 The problems associated with estimation in the presence of a random variable that lacks finite moments are well documented in the nonparametric estimation literature (see, for example, Scott [1992]).

Appendix: Impact of Bank Acquisitions

TABLE A1

Impact of Bank Acquisitions on Risk and Profitability[a]

	Risk					Profitability	
	S_A[b]	CV_A[c]	S_E[b]	CV_E[c]	Z	R_A[b]	R_E[b]
Acquired	−0.0011	–2.1600	−0.0186	0.7040	26.30	0.0008	0.0157
	7.94e–09	0.2940	1.91e−07	0.5640	0.00027	1.06e–05	4.24e−10
Acquirer	−0.0003	0.4150	−0.0036	0.5960	–11.10	0.0003	0.0085
	0.00116	0.2860	0.00604	0.1460	0.22400	0.00051	2.73e−11
Both–Both[d]	−0.0005	3.8400	−0.0054	0.7620	32.80	0.0005	0.0099
	7.56e−07	0.3850	9.71e−06	0.5240	0.00490	3.27e−07	3.87e−17
Both–Acquirer[e]	−0.0004	0.1530	−0.0046	0.2410	10.80	0.0001	0.0059
	5.59e−05	0.6940	0.0002	0.5110	0.3530	0.1430	1.18e−07

a. Compares the measures of risk computed eight quarters after the acquisition with the same measures computed eight quarters before.
b. Represents the standard deviation of the return on assets, S_A, and of the return on equity, S_E.
c. Shows the coefficient of variation of the return on assets, CV_A, and of the return on equity, CV_E.
d. Compares the postacquisition risk of the newly formed banking organization (Both) with the preacquisition risk of the hypothetical bank resulting from the sum of the acquired and the acquirer.
e. Compares the postacquisition risk of the newly formed banking organization (Both) with the preacquisition risk of the acquirer.
SOURCE: Authors' calculations.

TABLE A2

Risk Effects Based on the Covariance of Returns[a]

	Return on Assets	Return on Equity
8 x 8 Time Frame		
Covariance	2.22e–07	2.78e–05
	0.05850	0.6690

	Years after Acquisition							
	1	2	3	4	1	2	3	4
4 x 16 Time Frame								
Covariance	−6.03e–07	−1.29e–08	−1.12e–07	−1.37e–06	−9.56e–05	−1.98e–05	−1.98e–05	0.0002
	0.2710	0.9750	0.7630	0.1560	0.2370	0.7640	0.7350	0.1430

a. Compares the covariance between the returns of the acquired and the acquirer before and after the acquisition.
SOURCE: Authors' calculations.

TABLE A3

Impact of Bank Acquisitions on Risk[a]

	Years after Acquisition							
	1	2	3	4	1	2	3	4
	Acquired				Acquirer			
S_A	–0.0009	–0.0014	–0.0020	–0.0019	–0.0001	–0.0003	–0.0005	–0.0007
	0.00044	1.91e–06	4.06e–11	4.27e–10	0.27500	0.03300	4.59e–05	1.42e–05
CV_A	0.8950	1.4000	0.7060	0.2010	0.5060	0.8030	–0.0314	0.6810
	0.3120	0.09740	0.1540	0.63200	0.11700	0.11500	0.9380	0.1230
S_E	–0.0156	–0.0277	–0.0308	–0.0299	–0.0009	–0.0026	–0.0050	–0.0070
	0.00059	3.30e–06	4.77e–09	2.12e–08	0.51700	0.12200	0.00102	3.02e–05
CV_E	0.8490	1.2800	0.8760	0.5360	–0.8950	–0.4890	–1.790	–0.6070
	0.2690	0.0400	0.1910	0.33600	0.45000	0.6950	0.1610	0.6280
Z	20.200	51.100	85.700	66.000	–85.200	–65.30	–60.50	–28.00
	0.5460	0.0342	0.1010	0.00124	0.3010	0.4320	0.4610	0.7460
	Both–Both				Both–Acquirer			
S_A	–0.0003	–0.0005	–0.0007	–0.0008	–0.0002	–0.0004	–0.0006	–0.0007
	0.00931	–0.00027	8.36e–09	1.94e–08	0.1670	0.00573	1.03e–06	1.69e–06
CV_A	–0.5140	–1.0600	–0.2800	–0.6210	0.3260	–0.2280	0.5660	0.2300
	0.2170	0.0307	0.4890	0.1140	0.3810	0.6080	0.1090	0.48700
S_E	–0.0028	–0.0052	–0.0085	–0.0094	–0.0008	–0.0032	–0.0065	–0.0074
	0.03520	0.00139	2.09e–07	8.62e–08	0.5530	0.03380	1.09e-05	5.59e–06
CV_E	–0.5850	–0.7740	–0.1030	–0.3160	–1.360	–1.5500	–0.8830	–1.100
	0.2500	0.0439	0.7300	0.2160	0.2830	0.2020	0.4640	0.3590
Z	28.50	39.400	51.500	127.00	–73.40	–61.90	–51.80	24.300
	0.1040	0.0256	0.0155	0.00943	0.3790	0.4540	0.5330	0.8010

a. Compares the measures of risk computed four quarters before the acquisition with the same measures computed four quarters the acquisition, four quarters of the second year after the acquisition, and so on.

SOURCE: Authors' calculations.

TABLE A4

Impact of Bank Acquisitions on Profitability

	Years after Acquisition							
	1	2	3	4	1	2	3	4
	Acquired				Acquirer			
R_A	0.0007	0.0011	0.0017	0.0017	0.0002	0.0003	0.0005	0.0005
	0.00066	8.08e–06	3.67e–09	1.02e–08	0.06620	0.00133	4.67e–05	0.00028
R_E	0.0131	0.0224	0.0302	0.0309	0.0038	0.00791	0.0105	0.0123
	0.00028	4.74e–08	2.50e–11	3.76e–11	0.00016	6.98e–09	1.33e–11	7.35e–15
	Both–Both				Both–Acquirer			
R_A	0.0003	0.0006	0.0008	0.0008	–0.0001	0.0001	0.0001	0.0004
	0.0005	7.07e–07	6.88e–10	1.67e–08	0.25400	0.16500	0.00111	0.00773
R_E	0.0055	0.0109	0.0146	0.0160	0.0004	0.0058	0.0095	0.0108
	2.39e–06	2.37e–12	4.90e–17	3.40e–18	0.74000	8.52e–06	1.54e–10	1.95e–11

SOURCE: Authors' calculations.

References

Beatty, R.P., A.M. Santomero, and M.L. Smirlock. "Bank Merger Premiums: Analysis and Evidence," New York University, Salomon Brothers Center, Monograph Series in Finance and Economics, 1987–3, 1987.

Benston, G.J., W.C. Hunter, and L.D. Wall. "Motivations for Bank Mergers and Acquisitions: Enhancing the Deposit Insurance Put Option versus Earnings Diversification," *Journal of Money, Credit, and Banking*, vol. 27, no. 3 (August 1995), pp. 777–88.

Berger, A.N., W.C. Hunter, and S.G. Timme. "The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present, and Future," *Journal of Banking and Finance*, vol. 17, nos. 2–3 (April 1993), pp. 221–49.

Boyd, J.H., and S.L. Graham. "Investigating the Bank Consolidation Trend," Federal Reserve Bank of Minneapolis, *Quarterly Review*, Spring 1991, pp. 3–15.

Boyd, J.H., S.L. Graham, and R.S. Hewitt. "Bank Holding Company Mergers with Nonbank Financial Firms: Effects on the Risk of Failure," *Journal of Banking and Finance*, vol. 17, no. 1 (February 1993), pp. 43–63.

Clark, J.A. "Economies of Scale and Scope at Depository Financial Institutions: A Review of the Literature," Federal Reserve Bank of Kansas City, *Economic Review*, September/October 1988, pp. 16–33.

Craig, B.R., and J.C. Santos. "Performance and Asset Management Effects of Bank Acquisitions," Federal Reserve Bank of Cleveland, Working Paper No. 9619, December 1996.

Fraser, D.R., and J.W. Kolari. "Determinants of Small Bank Acquisition Premiums," in *Proceedings of a Conference on Bank Structure and Competition*, Federal Reserve Bank of Chicago, 1987, pp. 393–405.

Hobson, H.A., J.T. Maston, and J.T. Severiens. "Holding Company Acquisitions and Bank Performance: A Comparative Study," *Journal of Bank Research*, vol. 9, no. 2 (Summer 1978), pp. 116–20.

Hunter, W.C., and L.D. Wall. "Bank Merger Motivations: A Review of the Evidence and an Examination of Key Target Bank Characteristics," Federal Reserve Bank of Atlanta, *Economic Review*, September/October 1989, pp. 2–19.

Lawrence, R.J. "The Performance of Bank Holding Companies," Board of Governors of the Federal Reserve System, Staff Economic Study No. 55, 1967.

Litan, R.E. *What Should Banks Do?* Washington, D.C.: The Brookings Institution, 1987.

Meinster, D.R., and R.D. Johnson. "Bank Holding Company Diversification and the Risk of Capital Impairment," *Bell Journal of Economics*, vol. 10, no. 2 (Autumn 1979), pp. 683–94.

Mudur, U. "Economies of Scale and Scope in National and Global Banking Markets," in A. Steinherr, ed.,*The New European Financial Marketplace*. New York: New York University Press, 1992, pp. 31–48.

Rhoades, S.A. "Determinants of Premiums Paid in Bank Acquisitions," *Atlantic Economics Journal*, vol. 15 (March 1987), pp. 20–30.

———. "Bank Mergers and Industrywide Structure, 1980–94," Washington, D.C.: Board of Governors of the Federal Reserve System, Staff Studies, No. 169, January 1996.

Rose, P.S. "Profiles of U.S. Merging Banks and the Performance Outcomes and Motivations for Recent Mergers," in B.E. Gup, ed., *Bank Mergers: Current Issues and Perspectives*. Boston: Kluwer, 1989, pp. 3–28.

Santomero, A.M., and E.J. Chung. "Evidence in Support of Broader Bank Powers," *Financial Markets, Institutions & Instruments*, vol. 1, no. 1 (January 1992), pp. 1–69.

Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley and Sons, 1992, pp. 236–39.

Talley, S.H. "The Effect of Holding Company Acquisitions on Bank Performance," Board of Governors of the Federal Reserve System, Staff Economic Study No. 69, 1971.

Ware, R.F. "Performance of Banks Acquired by Multibank Holding Companies in Ohio," Federal Reserve Bank of Cleveland, *Economic Review*, March/April 1973, pp. 19–28.