Economic Review

Federal Reserve Bank of Atlanta

Economic Review

Economic Review

Federal Reserve Bank of Atlanta

The *Economic Review* of the Federal Reserve Bank of Atlanta presents analysis of economic and financial topics relevant to Federal Reserve policy. In a format accessible to the nonspecialist, the publication reflects the work of the Research Department. It is edited, designed, produced, and distributed through the Public Affairs Department.

Views expressed in the *Economic Review* are not necessarily those of this Bank or of the Federal Reserve System.

Material may be reprinted or abstracted if the *Review* and author are credited. Please provide the Bank's Public Affairs Department with a copy of any publication containing reprinted material.

Free subscriptions and limited additional copies are available from the Public Affairs Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713 (404/521-8020). Change-of-address notices and subscription cancellations should be sent directly to the Public Affairs Department. Please include the current mailing label as well as any new information. ISSN 0732-1813

Contents

Is federal government insurance of private pension plans, provided through the Pension Benefit Guaranty Corporation (PBGC), in danger of requiring bailout? This article focuses on private defined-benefit pensions that could wind up on the PBGC's balance sheet. The author gives an economic analysis of the agency's operations and the growing concern that taxpayer funds may be needed to fulfill its obligation as guarantor of private pension claims. The discussion points out similarities and differences between the problems with government pension insurance and the well-known problems with deposit insurance.

U.S. bank mergers are rising in both frequency and scale. Whether this development is a panacea or disaster for the banking industry is a subject of considerable debate. In this debate one important issue is that of postmerger cost savings. This article reviews several studies that investigate cost savings and reports the author's original research examining before-and-after performance of a sizable group of mergers of larger U.S. banks completed between 1982 and 1986. According to her findings, cost savings from mergers are far from automatic and shareholders and policymakers should carefully consider cost-savings claims case by case.

The authors provide a primer on path-dependent options, a new class of derivative financial instruments that is becoming increasingly popular. The article reviews the modern origins of path-dependent options, gives examples of reasons that investors find them attractive, and describes some essential features of three common types.

Cheating the Government: The Economics of Evasion
by Frank A. Cowell

Corporate Pensions and Government Insurance: Déjà Vu All Over Again?

Peter A. Abken

S ince the passage of the Employee Retirement Income Security Act (ERISA) in 1974, the federal government has insured part of the payments promised to retirees from private defined-benefit pensions.[1] ERISA established the Pension Benefit Guaranty Corporation (PBGC), a quasi-government agency, to provide pension insurance. The PBGC's board of directors comprises the secretaries of the U.S. Departments of Labor, the Treasury, and Commerce. The agency is intended to be self-financing, using premium income from the pension funds it insures as well as its own investment income and recoveries of assets from corporations that turn over pension plans to the PBGC. Those companies' terminations of defined-benefit plans because of insufficient assets would leave the PBGC with the responsibility for paying the pension claims.

This article focuses on the contingent liabilities of the PBGC—exposure to pension plans that could potentially wind up on the PBGC's balance sheet. As of year-end fiscal 1991 the agency had a deficit of $2.3 billion. More worrisome are the contingent liabilities, which have grown much faster than anticipated. The 1991 level of unfunded liabilities in the defined benefit pension system stood at $40 billion, $10 billion greater than the previous year (James B. Lockhart III 1991b, 3). The Office of Management and Budget (OMB) estimates the PBGC's contingent liability at the beginning of 1992 to be $43 billion (Budget 1992). Future PBGC income is projected to fall far short of this amount.

This article gives an economic analysis of the PBGC's operations and the growing concern that taxpayer funds may be needed to fulfill its obligation as guarantor of private pension claims. The article concludes with a discussion of the similarities and differences between the problems with government pension insurance and the well-known problems with bank deposit insurance.

The author is a senior economist in the financial section of the Atlanta Fed's research department.

The PBGC and Defined-Benefit Plans

The Pension Benefit Guaranty Corporation insures the pension benefits of approximately 40 million American workers. These workers are participants in approximately 85,000 defined-benefit retirement plans (PBGC 1992, 1). Some background is useful in delineating the PBGC's role in the pension system.

The crucial difference between defined-benefit and defined-contribution plans from the employer's vantage point is that the former can be underfunded whereas the latter is always, by definition, fully funded. To understand the distinction, it is helpful to consider the balance sheet of a corporate pension fund. This fund is operated as a trust fund separate from a corporation's other assets and liabilities, although the division is somewhat porous. A corporation's pension fund managers make investments in stocks, bonds, and other assets in order to pay for or fund its pension obligation, the liability item on the balance sheet. The actual value of future pension obligations is uncertain in the case of a defined-benefit plan because the promise is in terms of fixed payments to retirees at future dates, payments based on a number of variables—projected salary levels, employment levels, retirement ages, and life expectancies. The firm's ability to meet the obligation also depends on the expected rate of return on its investments, not just on its contributions over time. The higher the expected rate of return, the lower the necessary contributions to the fund. In contrast, no fixed benefits are promised through a defined-contribution plan, which makes the plan's beneficiaries bear all the risk for the investment results.

Before ERISA, participants in defined-benefit plans also faced the risk that a pension fund's assets would fail to meet the fund's pension obligations. In this case the current and future retirees would receive less than their promised benefits. The closing of the Studebaker automobile plant in South Bend, Indiana, in 1963 was a major incident contributing to the adoption of ERISA a decade later. Studebaker's defined-benefit plan was seriously underfunded, resulting in reduced or completely eliminated pension benefits to the automobile maker's 8,500 employees (Alicia H. Munnell 1982, 24; PBGC 1987, 22). Studebaker was not the only employer to terminate its pension plan during financial stress. The large drop in stock prices in 1973 and 1974 increased the risks for pension plan participants, as pension asset values tended to fall relative to liabilities. Concern about pension terminations resulting from economic risks and fraudulent management

of pension funds may have motivated the enactment of ERISA in 1974.[2]

A major rationale behind ERISA was to establish fiduciary standards for corporate pension plans. The basic premise is that pension plan sponsors are to manage their plans solely for the benefit of plan participants, the employees and retirees. ERISA prohibits a sponsor from investing substantial amounts of pension plan assets in the sponsor's own securities and requires the sponsor to eliminate underfunding of defined-benefit plans within a specified time period.[3] The act also spelled out employees' vesting rights in private pension plans, stipulating well-defined periods of employment after which the employee becomes legally entitled to the employer contributions to his or her pension.[4] Finally, as noted above, ERISA established the PBGC.

The PBGC has two separate insurance functions. First, it protects pension participants from loss of their pension benefits, up to prescribed guarantee levels, in the event of financial distress of the plan sponsor. This maximum benefit has periodically been revised upward as the Social Security wage base has increased. (On average, guaranteed benefits account for 90 to 95 percent of vested benefits; on average, about 90 percent of accrued benefits are vested.) In 1991 the PBGC's maximum guaranteed benefit for a participant in a single-employer plan who retires at age sixty-five with no survivor benefits was $2,250.00 per month (PBGC 1989, 16). Second, for pension plans that become trusteed to the PBGC, the agency pays a pension annuity to retirees (of no more than the maximum PBGC-guaranteed benefit level of $2,250.00 per month for each participant). The agency's annuity operation is a large part of its activity and consumes a major share of its resources (PBGC 1990a, 5).

As mandated by law, the PBGC runs two distinct insurance programs for single-employer and multiple-employer defined-benefit plans. The single-employer fund of the PBGC dwarfs the multiemployer fund in terms of its demands on the agency's financial resources. Multiemployer plans are arranged and administered by particular unions for their members. Employer contributions are negotiated through collective bargaining. For fiscal year 1991 benefits paid by the PBGC on trusteed single-employer plans were $514 million to 140,100 participants, whereas benefits paid on trusteed multiemployer plans during the same period were $2 million to 1,990 participants. Premium income to the PBGC for single-employer plans was $741 million as compared with $23 million from multiemployer plans (PBGC 1991c, 60).

The PBGC allows a corporate sponsor to discontinue a pension plan in two ways: standard termination and distress termination.[5] A standard termination does not involve the intervention of the PBGC. If a firm's pension plan has sufficient assets to meet its pension obligations, then a firm has the right to notify plan participants that it is terminating its defined-benefit pension plan and substituting another plan (for example, a defined-contribution plan) or otherwise paying off its obligation by making lump-sum payments to plan participants (usually then rolled over into Individual Retirement Accounts) or by buying annuity contracts from insurance companies for them. There are currently about 9,000 standard terminations of fully funded plans per year (Lockhart 1991c, 6).[6]

Pension plans often allow their corporate sponsors to dissolve a plan in the event a surplus (assets in excess of liabilities) builds up in a pension fund. This sort of plan discontinuation is known as a "reversion" because the surplus reverts to the firm. Reversions were common occurrences in the 1980s. In a study by the PBGC of 1,024 defined-benefit plans, 188 plans were terminated during the 1980-88 period, with about 80 percent of those plans overfunded at the date of termination (PBGC 1990b, 2).

Distress terminations can be initiated by either the PBGC or by a plan sponsor. The Pension Protection Act (PPA) of 1987 enumerated the following "distress criteria":

(1) Chapter 7 bankruptcy liquidation;
(2) Chapter 11 bankruptcy reorganization. (The employer must demonstrate to the court that liquidation would necessarily follow if the pension plan were not terminated.);
(3) A determination by PBGC that the employer is in such poor financial condition that, unless the plan is terminated, the employer cannot pay its debts when due and cannot continue in business;
(4) A determination by PBGC that, solely because of a decline in the employer's work force, pension costs have become "unreasonably burdensome." (PBGC 1989, 18-19)

For a plan sponsor to make a claim against the PBGC, one of these four statutory distress tests must be met.

In the event of a distress termination, the PBGC ensures that guaranteed benefits are paid to plan participants. If a firm has sufficient assets to meet the guaranteed benefits, then those assets are paid out to participants as in a standard termination. On the other hand, if the assets are insufficient, the PBGC makes up the shortfall out of its own assets, taking over the plan as its trustee. The agency assumes all of the liabilities of the pension fund as well as all assets.

In addition to receiving the assets and liabilities of a pension fund subjected to distress termination, the PBGC has a claim on up to 30 percent of the net "assets" of the firm. This provision of ERISA is intended to help the agency offset the unfunded portion of pension funds turned over to it. The meaning of net assets is ambiguous for reasons that will be better appreciated after further discussion.[7] As codified in ERISA, Congress intended net assets to mean corporate assets minus corporate liabilities, separate from the assets and liabilities within its pension fund. However, corporate equity, the market's valuation of net worth, in principle also includes the value of a firm's right—or option—to shed itself of its pension obligation by shifting it to the PBGC. The Single Employer Pension Plan Amendment Act (SEPPAA) of 1986 and Pension Protection Act of 1987 reduced the value of that option by limiting a plan sponsor's ability to terminate a pension plan voluntarily. In any case, there is no precise market measure (as opposed to accounting measure) of net corporate assets.[8]

In initiating a defined-benefit plan termination, the PBGC has the authority to place liens on the nonpension assets of solvent, ongoing firms to pay benefits if the pension fund lacks sufficient assets to meet those payments. In cases involving bankruptcy reorganizations, ERISA grants the PBGC priority status (equivalent to a federal tax lien imposed by the Internal Revenue Service) as a creditor in bankruptcy proceedings. A September 1991 U.S. District Court ruling regarding a particular bankruptcy case has denied the PBGC's priority status, a decision the agency is contesting. This development as well as the agency's current financial problems will be discussed below.

Pension Plan Terminations and the PBGC Deficit

Since the PBGC's creation in 1974 the agency has had to take over 1,600 pension plans involving 300,000 participants. The agency funds its obligations to trusteed plan participants through premium income from insured defined-benefit plans, through investment income from assets acquired in the course of taking over terminated pension plans, and through recoveries from firms terminating their pension plans.[9] Plan sponsors'

net worth in bankruptcy has not often been much greater than zero, resulting in recoveries in recent years that have averaged only 13 cents for every dollar of unfunded PBGC-guaranteed pension liabilities and unpaid pension contributions (PBGC 1992, 7). The PBGC also assesses interest charges and substantial penalties on plan sponsors who make late contributions to their pension funds or who underpay the minimum statutory contributions.

As Chart 1 shows, PBGC liabilities have exceeded its assets in every year of the PBGC's operation since its founding. The perennial surpluses accruing to the multiemployer fund have done little to defray the vastly larger perennial deficits of the single-employer fund. As of the end of the 1991 fiscal year (September 30), the PBGC's net deficit was $2.3 billion (PBGC 1992, 1). Currently, the agency has more than adequate cash flow to pay out pension benefits and meet operating expenses. In the long term, however, the agency's prospects are less sanguine. The biggest liability items on the PBGC's balance sheet are the present value of future benefits to participants in currently trusteed plans and net claims for "probable" terminations of defined-benefit plans. Even for plans currently

trusteed, determining the present value of future benefits is not straightforward because "a significant portion of the future benefits is based upon estimated liability" (PBGC 1990a, 41).[10] As in any similar calculation by a pension plan administrator, actuarial assumptions must be used that are subject to error and revision.

More problematic is the evaluation of net claims for probable terminations. The PBGC includes "probable" claims (a term defined by the Financial Accounting Standards Board [FASB]) on its balance sheet as a liability, even though a pension plan has not yet been trusteed to the PBGC. For example, the agency booked Pan Am pension plans in 1991 before the airline had filed for bankruptcy or sought termination of its pension plan (Lockhart 1991c, 2). The other FASB categories for future claims are "reasonably probable" and "remote."[11]

The PBGC has consistently underestimated future net claims. The problem is that there is no reliable statistical method for predicting bankruptcy of plan sponsors. A preponderance of the PBGC's liabilities have come from the terminations of a relatively few large pension plans concentrated in the automobile, steel,

Chart 1
PBGC Accumulated Deficit

Dollars in billions
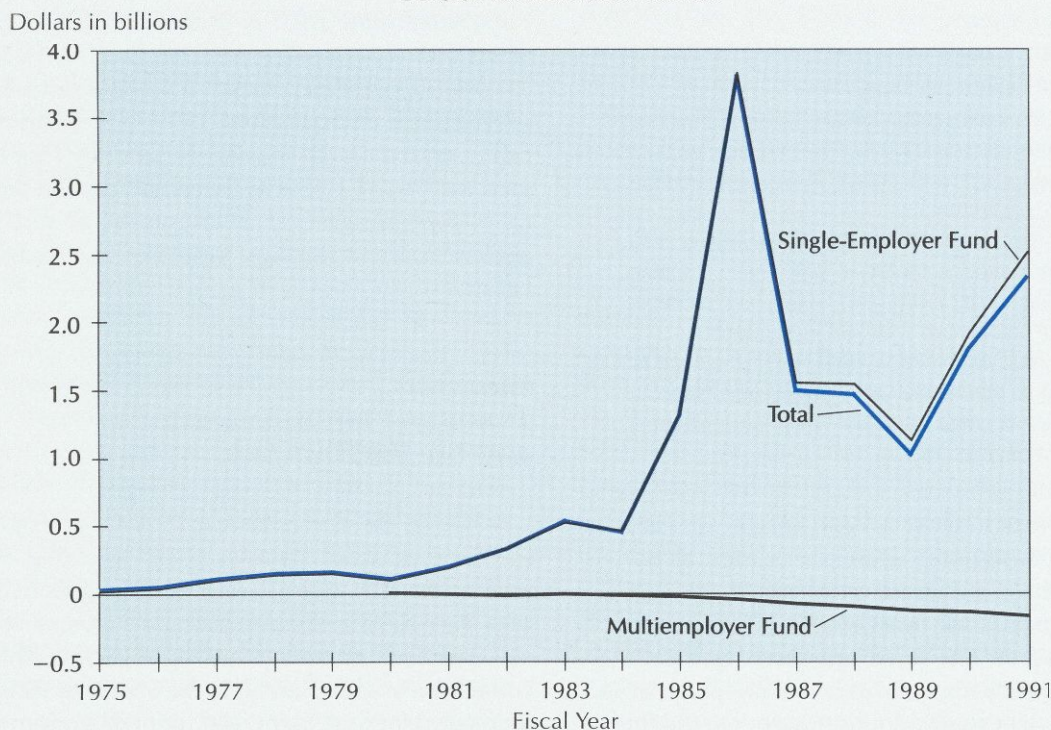
Source: PBGC (1991c) and Munnell (1982).

and airline industries (PBGC 1987, 36-37; PBGC 1991b, 2).

Table 1 shows the claims experience from single-employer plans during the PBGC's history as well as trends in those terminations. For the most recent period from 1986 to 1991, the number of terminated plans has declined compared with the previous two periods, but the net losses to the PBGC during the most recent six-year interval are almost two-and-a-half times greater than net losses during the agency's first eleven years' experience. The net losses from twenty-eight terminated plans realized in 1991 alone account for 25 percent of accumulated net losses since the PBGC was founded. Probable net losses identified in 1991 for fifteen plans amount to 20 percent of the cumulative net losses (PBGC 1991c, 6-7).

The second panel showing trends of terminated plans reveals that while the funding level (pension plan assets as a percentage of liabilities) has remained steady over time, the average net loss per terminated plan has risen markedly to $6.9 million. Recoveries by the PBGC in bankruptcy court have dropped sharply over the years to an average of 13 percent of net plan underfunding.

Among the major plan terminations of 1991, the largest was Pan American World Airways, with an unfunded liability of $900 million (including non-guaranteed benefits). The seven defined-benefit plans of Eastern Airlines, underfunded by $700 million, also significantly boosted the PBGC's accumulated deficit. Overshadowing both of these terminations is that of the LTV Corporation, which declared bankruptcy in 1986. One of its subsidiaries, LTV Steel, had four defined-benefit plans with $2.5 billion in underfunded liabilities at the time. The PBGC disputed three of the terminations and reinstated these plans, which resulted in litigation that eventually went to the Supreme Court.[12] In June 1990 the court upheld the PBGC's right to restore LTV's pension funds.

A September 1991 U.S. District Court decision (LTV 1991) threatens to terminate the largest of the LTV plans that the PBGC and Supreme Court restored. If the decision stands and ongoing bankruptcy reorganization negotiations fail, $1.6 billion in liabilities could be added in the near future to the PBGC's balance sheet from the most severely underfunded LTV pension plan. In February 1992 the LTV Corporation

Table 1
Claims Experience from Single-Employer Plans[a]

Year of Termination	Number of Plans	Benefit Liability	Trust Plan Assets	Recoveries from Employers	Net Losses
1975 - 1980	686	$ 570	$ 227	$ 65	$ 277
1981 - 1985	606	1,358	479	161	719
1986 - 1991	352	4,839	2,033	361	2,445
Total Terminated	1,644	6,768	2,739	587	3,442
Probable	15	1,552	595	181	776
Total	1,659	$8,319	$3,334	$768	$4,217

Trends of Terminated Plans[a]

Year of Termination	Funding Level (Percent)	Recoveries as a Percent of Net Underfunding	Average Net Loss Per Terminated Plan
1975 - 1980	40	19	$0.4
1981 - 1985	35	18	1.2
1986 - 1991	42	13	6.9

[a] Values shown are in millions of dollars. Stated amounts are subject to change until PBGC finalizes values for liabilities, assets, and recoveries.

Source: PBGC (1991c).

and its creditors (including the PBGC) reached an agreement on a revised reorganization plan, which was filed with the bankruptcy court in New York. LTV arranged a long-term funding plan with the PBGC to eliminate $3.2 billion in underfunding in the three disputed plans. Success of the reorganization plan hinges on union concessions, which are expected to be difficult to negotiate (Christi Harlan 1992). Meanwhile, both the PBGC and the Department of Labor have appeals pending concerning the September district court ruling.

Table 2 lists the PBGC's "Top Fifty" sponsors of underfunded pension plans, which the agency began publishing in 1991. The corporations are ranked by their funding ratio, the proportion of assets to guaranteed benefits liability. The LTV Corporation heads the list with a funding ratio of 6 percent and an unfunded guaranteed liability of $3.2 billion. General Motors is conspicuous for having the largest unfunded guaranteed liability of $7.1 billion, although its funding ratio stands at 83 percent. The automobile, steel, and airline industries are responsible for about 75 percent of the total underfunding among firms in the top fifty (PBGC 1991b). The PBGC's aggregate top-fifty liability level rose 50 percent from 1989 to 1990, from $14.2 billion to $21.5 billion. The cut-off level for underfunding to be included on this list rose from $32 million to $47 million over this period. Although this threshold is an arbitrary level, it does indicate a deterioration in funding of pension plans. (At the 1989 threshold, sixty-seven companies would have appeared on the 1990 list.)

According to the PBGC, there is about $40 billion in unfunded pension liabilities among all PBGC-insured defined-benefit plans, both single-employer ($31 billion) and multiemployer ($8 billion-$10 billion). By FASB category, the PBGC estimates $776 million in probable losses, $13 billion in reasonably possible losses, and $18 billion in remote losses for single-employer plans (PBGC 1991c, 10-11). These separate figures add up to the overall single-employer underfunding exposure. The PBGC analyzes the financial condition of plan sponsors individually and assigns them to these categories.

The most serious underfunding has arisen for so-called flat-benefit plans. These establish a monthly dollar benefit for each year of service and are designed for unionized hourly workers. Union contract negotiations, typically at three- to five-year intervals, tend to ratchet up retirement benefits before previously negotiated benefits are funded, increasing the degree of underfunding. The LTV plan mentioned above is such a flat benefit plan. In contrast, final salary plans generally anticipate salary increases and tend to be overfunded plans. The average final salary plan has a funding ratio of 145 percent, whereas the average flat benefit plan has a ratio of 75 percent (Lockhart 1991c, 3).

Most terminations of underfunded pension plans occur during times of financial stress for a sponsoring firm, when contributions are likely to be lowest. Contributions to a pension plan are a drain on corporate cash flow; thus, when pressed financially, firms will cut back on contributions to whatever extent they can. Although underpayment and late payment of required minimum contributions are subject to substantial penalties, sponsors may nonetheless elect to pay the penalties. They also have some discretion in choosing the actuarial assumptions—interest rate, retirement age, and so forth—that are used to calculate the minimum contribution. The present value of future pension benefits is particularly sensitive to choice of interest rate used for discounting future anticipated benefits. A higher interest rate reduces the present value of future benefits and in turn reduces the minimum contribution.

Differing assumptions can have dramatic effects on estimates of pension liabilities. As mentioned above, Eastern Airlines has turned over its pension plans with $700 million in unfunded liabilities to the PBGC. The executive director of the PBGC reported that estimates of Eastern's termination liability varied by as much as $400 million dollars, depending on the assumptions used to determine the discounted value of future pension benefits (Lockhart 1991c, 3).

Another factor that can allow a defined-benefit plan to become severely underfunded is the minimum funding standards established by ERISA. The amortization schedule permitted for certain liabilities can be longer than the time horizon over which actual benefits are paid out to beneficiaries, inducing net cash outflows from a fund (PBGC 1987, 26-27). Faster amortization schedules established by the Pension Protection Act of 1987 have partially rectified this problem (see Richard A. Ippolito 1989, 135-39).

Financial distress that leads to layoffs often saddles the PBGC with additional liabilities. Plant closings often entail "shutdown benefits," which essentially are early retirement benefits usually not funded by an employer in advance. These benefits are guaranteed by the PBGC. Minimum funding standards have not dealt adequately with such contingencies (PBGC 1987, 27). Because of incentives to underfund and because of increased rates of early retirement, funding ratios tend

Table 2
Top Fifty Companies with the Largest Underfunded Benefit Liability Guaranteed by PBGC[a]

Company	Guaranteed Benefits Liability[b]	Assets[b]	Unfunded Guaranteed Liability[b]	Funding Ratio (Percent)
LTV	$ 3,387	$ 204	$ 3,183	6.0
Jesup Group	132	27	106	20.3
CF&I Steel	219	55	164	25.0
Sharon Steel	213	64	150	29.9
Blaw-Knox	106	36	70	34.2
Loews	239	101	139	42.0
Keystone Consolidated Industries	178	79	99	44.4
American National Can	816	387	429	47.4
Uniroyal-Goodrich	826	392	434	47.5
Borg-Warner	200	98	102	49.1
New Valley	640	336	304	52.4
Harnischfeger Industries	161	86	75	53.2
Bridgestone-Firestone	510	279	231	54.7
National Intergroup	648	378	270	58.3
ACE Industries	149	89	59	60.0
Chrysler	8,415	5,087	3,328	60.5
Tenneco	368	224	144	60.8
Paine Webber Group	125	78	47	62.1
American Financial	150	93	56	62.3
Carter Hawley Hale Stores	158	98	59	62.3
Budd	490	312	178	63.7
Phelps-Dodge	167	116	51	69.3
Cyclops Industries	341	238	103	69.7
Bethlehem Steel	4,555	3,254	1,301	71.4
Rockwell International	643	461	182	71.8
White Consolidated Industries	227	170	57	74.9
Reynolds Metal	542	416	126	76.8
James River	263	204	59	77.5
Navistar International	2,481	1,945	536	78.4
UAL	431	344	87	79.7
Varity	284	227	57	80.0
Allegheny Ludlum	361	292	69	80.8
Northwest Airlines	432	355	78	82.0
Goodrich (BF)	533	441	92	82.7
General Motors	42,034	34,889	7,145	83.0
Maxxam	682	567	114	83.2
Raytheon	514	428	86	83.3
Weyerhaeuser	453	379	73	83.8
Westinghouse Electric	4,327	3,635	692	84.0
Honeywell	379	319	60	84.2
Kimberly-Clark	341	287	54	84.2
Trans World Airlines	1,251	1,061	190	84.8
Dana	533	458	76	85.8
Goodyear Tire & Rubber	1,083	933	150	86.2
Deere & Co	1,039	911	128	87.7
Armco Steel LP	622	556	66	89.4
Allied Signal	1,287	1,199	88	93.1
Armco	797	743	54	93.2
Textron	730	681	49	93.3
National Steel	829	779	50	94.0
Totals	$86,287	$64,788	$21,499	75.1

[a] Underfunding in pension plans guaranteed by PBGC is a measure of exposure for the pension insurance program. Experience has indicated that the PBGC's losses after a plan terminates often exceed these estimates because of lower contributions and greater-than-anticipated early retirements. In addition, detailed information about a plan's participants and provisions allow more refined valuation methods to be used, which often produce higher liability estimates; for example, such a valuation of TWA's plans done in August 1991 produced a $440 million estimate of underfunding for PBGC insured benefits.

The list does not include nonguaranteed benefits or potential recoveries by the PBGC upon termination. This list implies no rating or statement of financial condition of the companies. Liabilities have been adjusted to PBGC's assumptions at time of evaluation (7.25 percent and UP-84 mortality table), and excludes, whenever identified to the PBGC, plans sponsored outside the U.S. and domestic nonqualified plans. Totals may not add due to rounding.

[b] Values shown are in millions of dollars.

Source: PBGC, based on Standard & Poor's Compustat Services, Inc., of 1990 annual report data, 1988 IRS Form 5500, 1990 PBGC Premium Payment files, and information provided by contacted companies.

to deteriorate rapidly for plan sponsors on the brink of bankruptcy. The average underfunded plan that eventually terminates has a funding ratio that declines from 80 percent five years before termination to 40 percent upon the date of termination (PBGC 1991c, 11).

The Top Fifty list is a crude indication of the PBGC's exposure to underfunded plans. Many of the sponsors of underfunded plans on the list may not go bankrupt; some eventually may, and the degree of underfunding, as mentioned, tends to increase dramatically before termination. Some sponsors of currently fully or overfunded plans may also become financially distressed and underfunded in the future. Such plans also contribute to the PBGC's exposure.

A potentially better measure of the PBGC's contingent liabilities is available using options pricing methods. This approach also helps illuminate the nature of the claims against the agency.

An Options-Based Analysis of PBGC Contingent Liabilities

Before the enactment of ERISA, receiving promised pension benefits was not a sure thing. Payment of these benefits depended on the corporate pension fund's financial soundness and ultimately on the soundness of the sponsoring firm. Such contingencies lend themselves to analysis in terms of options, also often referred to as contingent claims. Pension benefits both before and after ERISA are types of contingent claims.

One way to assess a pension plan participant's claim is to determine what it is worth upon immediate termination of a plan. Before ERISA, if pension assets exceeded vested benefits, participants (employees and retirees) would receive the benefits promised. If assets fell short of vested benefits, then the most that could be received would be the value of the assets. The pension promise therefore depends on the relative magnitude of pension assets, denoted by A, and vested benefits, denoted by B. Symbolically, the payoff to plan participants upon termination of a plan is $min(A, B)$, that is, the minimum or lesser of A or B.

Pre-ERISA defined-benefit plans placed the risk of pension plan performance on the shoulders of the plan participants. Underfunding or weak pension asset returns could result in less than their promised benefits being received. If the pension fund experienced a surplus ($A - B > 0$), the plan sponsor reaped the gain; plan participants were entitled to no more than their promised benefits.[13] On the other hand, if the plan suf-

fered a deficit ($A - B < 0$), the plan sponsor had no liability for the shortfall and in fact had the right to terminate underfunded plans at its discretion (Gregory R. Niehaus 1990, 56).[14] The plan participants faced the downside risk.

The terminal payoff of a defined-benefit plan can be cast in terms of a standard options-based analysis. An option contract conveys the right, but not the obligation, to buy (call option) or sell (put option) an underlying asset at a specified price (the strike or exercise price) at a future date. Only the payoffs on options at the terminal date are considered at this point. Viewing the analysis in terms of a terminal date is an abstraction because pension obligations do not fall due at a single date. Nevertheless, the simplification is instructive, and the more realistic situation can be handled in a similar manner.[15] As noted above, the terminal payoff to pension plan participants is $min (A, B)$. Their exposure to the pension surplus (or deficit if negative) $A - B$ may be written as $min (A, B) - B$ because if $A \geq B$ the surplus to the participants is $B - B = 0$ and if $A < B$ the deficit borne by the participants is $A - B < 0$. After some simple algebraic manipulations, one can show that

$$min(A, B) - B = -max(0, B - A).[16]$$

The right-hand side is the negative of the payoff for a put option with strike price B, the promised benefit. The negative of the payoff is received for a short position in the put; that is, plan participants have effectively sold a put to the plan sponsor by taking part in the pension plan. The plan sponsor can "put" any deficit to the participants while retaining the right to any surplus. That right or option to the surplus is simply a call option with strike price equal to B, the promised benefits: $max(0, A - B)$. The division of the pension surplus or deficit is therefore

$$A - B = max(0, A - B) - max(0, B - A),$$

where the first term on the right-hand side is the plan sponsor's call option and the second term is the plan participants' short put option. (The equality can be readily verified by considering the cases $A \geq B$ and $A < B$.)

The decomposition of the pension surplus also holds any time before maturity of the pension obligation by appealing to the standard put-call parity relationship (Robert C. Merton 1973).[17] The parity relationship is in terms of the present values of the call and put payoffs and the present values of the pension assets and liabilities. The present values of the call and put payoffs are,

in fact, option prices, which can be determined using option pricing techniques. Like any other options, these pension options are particularly sensitive to the volatility of the rate of return on the underlying assets—here, pension assets, which will increase the value of both the call and the put, an undesirable outcome from the perspective of pension plan participants.

At first glance, it might appear that workers were subject to being victimized in the event their pension plans were terminated. However, William F. Sharpe (1976) points out that in the pre-ERISA period employees' labor contract negotiations would in principle take into account corporate pension-funding decisions. The overall employee compensation would be the sum of the present value of the promised pension benefits and the present value of current and future wages minus the value of the pension put (which employees have effectively granted the employer). Contract negotiations could offset employer actions on the pension plan—particularly those that would increase the value of the corporation's put on pension assets. In other words, employees could implicitly force the employer to pay compensation (such as raising wages) for the value of the pension put that the employer holds.

Ippolito (1985) takes a provocative, controversial view of worker-employer pension relationships. The crux of his argument is that prior to ERISA union pension plans tended to be underfunded by firms in order to constrain union bargaining power. Excessive wage demands or other demands that would reduce productivity would be limited because union workers became claimants on their employers through the pension plan. They effectively held a long-term "pension bond," represented by the underfunding, that they would forfeit if their employer went out of business. In other words, union demands that would increase the likelihood of bankruptcy would also increase the chance that workers would fail to receive their full pension benefits.

After enactment of ERISA, the PBGC assumed most of the risk of the underfunding that pension plan participants had formerly borne. The reasons behind ERISA's enactment may not simply have been a matter of protecting workers' pension claims. Ippolito argues that "[t]he evidence is overwhelming that ERISA was not, and is not, a public-interest piece of legislation. The data suggest that the regulation was enacted to benefit a small, concentrated group: union workers in dying firms. This is done through a government agency [the PBGC]. Claimants essentially can determine the timing and magnitude of their transfers, and

they pay nothing for the privilege. All defined benefit plan participants are taxed a small amount to pay for the transfers" (1988, 120). In other words, by guaranteeing a large portion of pension benefits, ERISA enabled unionized workers to extract greater rents (that is, higher wages) from their employers. While premiums that fund the PBGC are assessed against all defined-benefit pension plans, the principal beneficiaries of pension insurance are unionized workers participating in underfunded plans.[18]

The post-ERISA analysis of pension claims is similar to the pre-ERISA case. A central statutory provision of ERISA is that, in the event of a plan termination, the PBGC has a priority claim on 30 percent of the plan sponsor's net worth.[19] Thus, the strike price for the pension put is the same as before, equal to vested benefits B, but the underlying asset for the put becomes the pension assets plus 30 percent of net worth, denoted as NW. The pension surplus (or deficit) accruing to the plan sponsor upon termination becomes

$$A - min(B, A + .3 \cdot NW).$$

This statement may be transformed algebraically to

$$A - B + max[0, B - (A + .3 \cdot NW)].$$

The term involving the *max* operator is the PBGC pension put, which has only positive value to the sponsor. The interpretation of either of these equivalent expressions is that for a sufficiently underfunded plan ($A + .3 \cdot NW < B$) the pension deficit is no greater than 30 percent of net worth: the plan sponsor terminates the pension and turns over the plan's assets plus 30 percent of net worth (which is often zero in bankruptcy) as well as the plan's liabilities to the PBGC. Otherwise, for $A + .3 \cdot NW \geq B$ the pension deficit (or surplus) is $A - B$. An overfunded plan ($A - B > 0$) never winds up on the PBGC's balance sheet. However, the sponsor may terminate an overfunded plan to extract its surplus, as mentioned above.

In principle the PBGC should base its premiums for pension insurance on the value of the sponsor's pension put. Doing so would eliminate any incentive for the firm to exploit the pension insurance. Increasing the riskiness of pension assets (for example, buying volatile—high beta—stocks or junk bonds) would increase the value of the pension put, but higher premium assessments would negate any benefit, as Sharpe (1976) has argued. Otherwise, "[t]he corporation [plan sponsor] should adopt those policies which maximize the difference between the market value of

the option represented by the insurance contract and the premium charged for the contract" (1976, 193).

Empirical Evidence

There is a dearth of empirical work on corporate pension funding policy. The available evidence gives only weak support to Sharpe's (1976) proposition that plan sponsors would exploit their pension insurance.

The pricing of the insurance mandated by Congress clearly made it susceptible to moral hazard. Indeed, from the PBGC's inception until 1988, plan sponsors paid only a flat-rate premium for their pension insurance. Randolph Westerfield and William Marshall (1983) estimate a positive correlation between the proportion of equity investment and the degree of pension plan underfunding after enactment of ERISA. This calculation would be consistent with maximizing behavior because a greater share of equity would tend to increase the volatility of pension assets and thereby increase the pension put's value. However, the correlation is statistically insignificant. Zvi Bodie and others (1987) observe that higher-risk plan sponsors held riskier portfolios in their pension funds, but this situation occurred for only three firms in a sample of 539, and two of the three subsequently shifted their pension portfolios to 100 percent bond investments while financially distressed.

On the other hand, Niehaus (1990) found that, following the enactment of ERISA and PBGC insurance, promised pension benefits increased for single-employer plans covered by pension insurance compared with other plans that were not. (PBGC insurance of multiemployer plans was not instituted until 1980, and some single-employer plans had benefit levels that exceeded guaranteed levels.) Niehaus interprets this evidence as being consistent with the existence of a moral hazard problem; that is, single-employer plan sponsors could exploit the then-prevailing flat-rate insurance premium by negotiating higher promised pension benefits for employees while reducing nonpension benefits and wages. These actions would effectively increase the value of the pension put. With PBGC insurance coverage, employees would have no incentive to worry about underfunding of their pension claims if they were at or below guaranteed levels. Niehaus's statistical tests indicate that the moral hazard was greatest for pension plans covering union workers, consistent with Ippolito's (1988) contention.

Reforms of PBGC Insurance

In response to the mounting PBGC deficit accrued since the agency's early years, legislation has been passed, with a lag, to raise the premium for pension insurance. The original premium was $1 per plan participant per year. A major flat-rate increase was imposed by SEPPAA in 1986, raising the rate from $2.60 to $8.50, which at the time was still inadequate to amortize the PBGC's deficit (PBGC 1987, 24). The Pension Protection Act of 1987 increased the basic flat-rate premium to $16 per participant and instituted a variable-rate premium for underfunded plans that increased incrementally with the degree of underfunding. The variable-rate premium was capped at $34 per participant (PBGC 1988, 6). The current flat-rate premium is $19 per participant, with a maximum flat and variable premium of $72 per participant (Lockhart 1991b, 5).

A variable-rate premium goes some way toward assessing a risk-based premium to the extent that underfunding is associated with financial distress. However, this rate structure is still subject to moral hazards, as discussed by James E. Pesando (1982) and reiterated by Jack L. VanDerhei (1990). For example, if a plan sponsor issues bonds to finance contributions that reduce underfunding, the value of the pension put does not necessarily fall, but the probability of bankruptcy rises. The sponsor is free to reduce underfunding by purchasing risky assets like common stocks that tend to increase the value of their pension put with no effect on their variable-rate premium.

The variable-rate premium and distress tests for plan termination were enacted to put PBGC finances on a sounder footing and reduce the cross-subsidization of weak pension plans by strong plans. Despite these reforms, the PBGC faces both immediate and longer-term financial challenges. By imposing distress tests, SEPPAA and PPA greatly limited a plan sponsor's ability to shed its pension obligations voluntarily—in other words, exercise its pension put—while the sponsor is financially healthy. However, most major terminations occur during bankruptcy proceedings. The LTV bankruptcy case poses a particularly serious threat to the PBGC's finances. The U.S. District Court ruling in September 1991 included three detrimental actions: (1) it allows bankrupt firms undergoing reorganization to suspend pension plan contributions while permitting pension liabilities to increase during bankruptcy proceedings; (2) it eliminates the PBGC's priority status in bankruptcy—the agency would be on a par with other unsecured creditors; and (3) it denies

the PBGC the right to specify the actuarial assumptions used to assess their claims against a sponsor (Lockhart 1991c, 5). The ruling turned upon the primacy accorded bankruptcy law relative to pension law in bankruptcy proceedings.

The PBGC is seeking the following legislative changes to ERISA and to the Bankruptcy Code:

(1) Clarify that [the PBGC has] priority claims for most unpaid pension contributions and for underfunding up to 30 percent of the net worth of the controlled group [plan sponsor(s)];

(2) Change the priority claim for underfunding to the greater of 30 percent of net worth (as under current law) or a small, but gradually increasing percentage of underfunding;

(3) Give tax priority to claims for underfunding resulting from shutdown benefits triggered within three years of termination because these heavily subsidized retirement benefits are not generally prefunded and are very costly to the insurance fund—approximately half a billion dollars to date; and

(4) Give PBGC the option to be a member of creditors' committees so that [the agency] can have access to information routinely available to other creditors and help speed reorganizations. (Lockhart 1991c, 5)

These central proposals contained in the Pension Protection in Bankruptcy Act of 1991 were submitted to Congress in November 1991. For item (2), the new bill stipulates that the priority claim be the greater of 30 percent of net worth or a percentage of underfunding that begins at 10 percent and rises in 2 percentage point annual increments to a maximum of 50 percent in 2012.

The PBGC's longer-term challenge is to its solvency and capacity to remain self-financing. According to the agency's ten-year forecasts, its current premium schedule will fail to narrow the fiscal year 1991 deficit, even if its projection is based on the most optimistic scenario—that the agency will experience annual net losses equal to the average realized since 1974. Under these conditions, the deficit would widen slightly to $2.7 billion by the end of fiscal year 2001. The PBGC's most pessimistic scenario, which includes all "reasonably possible" losses, puts the deficit at $17.9 billion at the end of the forecast horizon, up from the $11.4 billion forecast a year ago (PBGC 1991c, 12).

The Office of Management and Budget has measured the PBGC deficit using an options pricing methodology, intended for broad application as a means to value the contingent liabilities of federal insurance programs like pension and deposit insurance. This approach estimates the value of the pension put for each insured plan and then aggregates those values to obtain the PBGC's contingent liability; it is similar to that used in Alan J. Marcus (1987) and Arturo Estrella and Beverly Hirtle (1989) (see Budget 1992, chapter 13). In other words, the contingent liability is the present value of future claims on the PBGC. Use of this methodology is an important part of a shift from cash accounting to accrual accounting proposed in the Bush Administration's Economic Growth and Job Creation Act of 1992, which includes a number of provisions to reform pension insurance.[20]

The OMB estimates that the PBGC's contingent liability at the beginning of 1992 was $43 billion for single-employer plans (Budget 1992, part 1, 277). (The much smaller contingent liability for multiemployer plans has not been estimated thus far.) The estimated present value of the PBGC's premium income is about $9 billion, and the estimated accrued cost of pre-1992 liabilities is $6 billion, reflecting the unfunded costs of past plan terminations. By these measures the overall underfunding of the PBGC is therefore approximately $40 billion. Although the computation of the contingent liability is subject to error because it is based on a particular options pricing model as well as on many simplifying assumptions, it is reasonable to conclude that the underfunding is sizable. For the PBGC to remain self-financing in the long run, the agency's underfunding gap needs to narrow over time. Without reform of the pension insurance system, taxpayer involvement may be needed to satisfy future pension claims on the PBGC.

The Bush Administration's proposed pension insurance reforms, besides including the bankruptcy reforms already submitted to Congress, provide for strengthened minimum funding requirements and a freeze on additional pension guarantees for chronically underfunded plans. Underfunding would be amortized faster under the proposed law, and new pension benefit promises appended to longstanding underfunded plans would not receive PBGC insurance coverage. The OMB estimates that without these reforms the PBGC deficit on an accrual basis would be $20 billion by 1997. With enactment of these reforms, the OMB projects a PBGC surplus of $4 billion (Budget 1992, part 1, 276).

Alternative measures could also shore up PBGC finances. These could include reducing PBGC exposure—and potential taxpayer exposure—by scaling back the

amount of guaranteed pension benefits. As discussed above, workers can be compensated for increased exposure to a higher "deductible." At the same time, a higher deductible improves the incentive to limit underfunding. However, similar proposals to reduce the amount of deposit insurance proved to be politically unacceptable. Reducing pension guarantees would undoubtedly also be contentious. In fact, there is a constituency for greatly expanding the federal safety net for pensions.[21]

The most radical (and politically unrealistic) course of action would be to privatize pension insurance. Ippolito (1986) has proposed a system of private insurance backed up by federal government reinsurance. The premium structure for pension insurance would be established in the private market, though the federal government would mandate the level of insurance coverage. In turn, private insurers would insure themselves against catastrophe (a deep depression and corresponding collapse of asset values) by buying insurance from the federal government. The advantage of this approach is that private insurers would have the economic incentive not to underprice the value of the pension puts held by the insured plan sponsors.

Deposit Insurance and Pension Insurance

The financial problems of the PBGC are relatively obscure compared with those of the Federal Deposit Insurance Corporation (FDIC) and the now defunct Federal Savings and Loan Insurance Corporation (FSLIC), whose role as deposit insurance agency has been assumed by the Savings Association Insurance Fund (SAIF). Since the bailout of savings and loans (S&Ls) began two years ago, $80 billion has been spent toward the "resolution" of bankrupt thrifts. The Bush administration is requesting another $80 billion to continue the ongoing operations of the Resolution Trust Corporation, which is in charge of liquidating problem thrifts (Stephen Labaton 1991). The banking bill passed on November 27, 1991, authorizes the FDIC to borrow $30 billion from the Treasury to recapitalize its depleted Bank Insurance Fund and grants an additional $40 billion credit line with the Federal Financing Bank to meet working capital needs (Robert M. Garsson 1991). Costs are escalating partly because of the government's inefficiency in managing the resolution process (Edward J. Kane 1990).

A recent comparison of federal insurance programs indicates the magnitude of money involved for deposit insurance and pension insurance. Using its new accrual budgeting methods, the OMB projects that by 1997, under current law, the FDIC's Bank Insurance Fund will run a deficit of about $39 billion, the Resolution Trust Corporation's SAIF a deficit of $60 billion, and the PBGC's single-employer fund a deficit of $20 billion (Budget 1992, part 1, 276).

There is a close parallel between government insurance of bank and S&L deposits and pension benefits. Deposit insurance has also been analyzed in terms of an option pricing framework. Important articles in this literature include Merton (1977), Marcus and Israel Shaked (1984), and Ehud I. Ronn and Avinash K. Verma (1986). De facto, the deposit insurance agencies have usually insured all depositors in the event that a depository institution (hereafter simply referred to as a bank) fails. The insured debt obligation of a bank is a deposit with which a bank acquires assets—loans and securities. Merton notes that although bank deposits have no fixed maturity, the term of deposit insurance could be regarded as the time between bank examinations. The value of deposits plus interest at the examination date is represented by D; the value of the bank's assets by A. If examiners find that a bank is insolvent $(A - D < 0)$, they can liquidate it or merge it with a solvent institution. In the event of insolvency, the deposit insurance agency, such as the FDIC, must pay the depositors the difference between the value of their deposits and the value of the bank's assets. Deposit insurance is effectively a put option held by the bank, giving it the right to sell its assets for D dollars to the FDIC. The terminal payoff of this put is $max(0, D - A)$; the put is "in the money" and therefore exercised by bank management when the bank is insolvent, $A < D$.

The deposit insurance put is an asset on a bank's balance sheet, just as the pension put is an asset on a pension plan sponsor's balance sheet. Appropriately priced deposit insurance (premium assessments equal in present value to the put) has no effect on the value of a bank. However, mispriced deposit insurance leaves the deposit insurance agency, and ultimately taxpayers, vulnerable to the moral hazard that bank managers will exploit the insurance. Expected returns on the assets can be increased by making riskier loans or buying riskier securities, with the downside risk borne by the FDIC. The increased value of the put (with no change in its cost) on the balance sheet boosts a bank's total asset value and its stock price. In the absence of insurance, depositors would demand a

risk premium on their deposits as compensation for default risk. Appropriately priced deposit insurance would make bank managers pay an insurance premium at least equal to the expected value of the losses to the deposit insurer.

Although the nature of the insurance guarantee and the incentive to exploit mispriced insurance is similar between pension fund management and banking, actual experience has been markedly different. As noted earlier, there is no significant evidence that pension funds allocate their assets to maximize their pension puts. In contrast, S&Ls and, to a lesser extent, commercial banks have taken advantage of deposit insurance in a way that increases the value of their put options. For example, many S&Ls were notorious for attracting deposits through high deposit rates and use of brokered deposits, all of which were insured, to expand assets rapidly, especially by making high-risk loans to real estate developers and others. Depositors faced no risk and the money flowed in seeking the higher rates. The real estate collapse in the Southwest, following the plunge in oil prices in 1985, precipitated much of the current problems in the S&L industry.

Commercial bank portfolios have also become riskier in the 1980s, though the increased riskiness of assets on their balance sheet, by and large, probably reflected the changing characteristics of commercial banks' clientele more than an interest in benefiting from deposit insurance. During the 1980s their best corporate customers continued the trend of borrowing directly in financial markets, like the commercial paper market, rather than through commercial banks. Regulatory and legal restrictions also limit depository institutions' ability to diversify geographically and commercially. Again, banks' and thrifts' lending to firms in the oil industry is now the textbook example of this sort of risk. Thus, part of the problem with troubled depository institutions did not have to do with deliberate, reckless risk-taking but instead with unhedged exposure to local or regional economic adversity.

Pension funds are much better diversified. Weak cash flow from the plan sponsor can be offset by better investment performance of the pension assets. For example, funds may choose investments in foreign securities to hedge the economic risks of a deterioration in the domestic economy and in the sponsor's line(s) of business (Robert A. Haugen 1989). The corporate sponsor may also be well diversified in its lines of business. Plan sponsors generally operate with fewer restrictions under ERISA than commercial banks and other depository institutions do under their applicable banking laws.

Most of the major losses to the PBGC have been concentrated in a few large industries undergoing contraction and consolidation—as mentioned earlier, the automobile, steel, and airline industries. The major losses and overhanging contingent liability of the PBGC have also been concentrated in terminated flat-benefit plans for unionized hourly workers, particularly those in the three industries identified. The losses that have occurred in the pension system have resulted from funding decisions by plan sponsors, not from a deterioration in the assets backing the pension liabilities. Cash flow problems created by poor sales or lagging productivity, for example, force plan sponsors to curtail pension contributions and accelerate retirements. These problems contrast with those typically afflicting failed or failing depository institutions. Their problems stem from erosion of asset values, often on loans related to real estate. Once the FDIC (or Resolution Trust Corporation—and formerly FSLIC—for S&Ls) intervenes, the problems do not end, especially if the institution is allowed to continue operating (because the FDIC lacks the "working capital" to completely take over all troubled institutions). James R. Barth, Philip F. Bartholomew, and Michael G. Bradley (1990) found that the main factor generating losses in the process of thrift resolutions is the amount of time insolvent thrifts were allowed to operate before being liquidated or merged with another institution.

The assets that the PBGC receives in the course of a plan termination are generally highly marketable, unlike the illiquid and often nonperforming loans and mortgages and deteriorating foreclosed properties that deposit insurers take over. Losses continue to mount on these assets, often at an accelerating rate once in the hands of the deposit insurer (Kane 1990). The losses for pension funds accumulate before the PBGC intervenes, largely because of statutory restrictions on its actions. The agency is an involuntary creditor; it is obligated to provide insurance up to guaranteed-benefit levels, with no power to alter the terms of its insurance or influence the actions of pension plan sponsors.

Periodic escalation of pension insurance premium rates is not a viable strategy for maintaining solvency of the PBGC. The PBGC faces a more acute adverse selection problem than do deposit insurers—that is, higher premium rates will drive well-funded, well-managed pension plans out of the system. Deposit-taking is a core function of the banking business. Banks and

other depository institutions can relinquish their banking charters (and become nondeposit financial intermediaries) if deposit insurance makes them unprofitable, but this action is clearly drastic. In contrast, corporations sponsoring defined-benefit plans for their employees have the option of discontinuing their plans and substituting defined-contribution plans (or 401[k] plans) and thereby avoid paying PBGC premiums. This is a comparatively straightforward decision that would have little, if any, impact on the rest of the corporation. Standard terminations are commonplace events. Consequently, the PBGC, through congressional legislation, has much less latitude to assess high premiums against healthy pension plans in order to subsidize weak, underfunded ones than the FDIC does in making strong banks subsidize marginal banks.

Conclusion

After the well-publicized deposit insurance difficulties, the problems with pension guarantees may seem like a case of déjà vu, but in fact they have been longstanding, unresolved issues for the PBGC and policymakers. Many of the largest troubled "Top Fifty" pension funds have been underfunded for at least a decade (see Munnell 1982, 41-42), and PBGC net-worth deficits have persisted since its creation. The biggest questions are whether legislators have learned from the deposit insurance debacle and whether they will recognize and remedy the problems with pension insurance before it too requires a tax-payer bailout.

Notes

1. Most U.S. corporations offer retirement plans of two basic types to their employees: defined-benefit plans and defined-contribution plans. A defined-benefit plan promises a fixed nominal income upon retirement based on years of service and salary history. An integrated defined-benefit plan establishes retirement benefits that target the combined payments from Social Security as well as from the private pension plan. A defined-contribution plan provides for periodic contributions to a retirement account by the employer but does not guarantee the rate of return on the investment. The employee usually has a choice regarding the type of investment made on his or her behalf, typically allocating contributions among alternative investments in stock mutual funds, bond mutual funds, money market funds, or combinations of these. The defined-benefit pension plans nationwide hold assets of more than $1 trillion dollars (Lockhart 1991a, 5), which currently include more than 15 percent of the market value of corporate equities and stock mutual funds (Board of Governors 1991).

2. Ippolito (1988) argues that ERISA is a mechanism to transfer wealth from nonunionized workers (through their pension funds) to unionized workers. He finds little evidence to support the contention that ERISA was designed to prevent pension fraud or to insure against risk. His hypothesis is discussed below.

3. The amortization periods were substantially shortened under the Pension Protection Act of 1987. Ippolito (1989, chapter 8) contains a thorough discussion of the provision of this act.

4. Vesting for defined-benefit and defined-contribution plans must occur at least as fast as the following vesting schedules: Cliff vesting requires that a participant be fully vested no later than after five years of service (ten years for multiemployer plans). Graded vesting requires that a participant

be at least 20 percent vested after three years of employment and accumulate at least 20 percent vesting for each of the next four years. The employee must be fully vested by the end of seven years. Furthermore, employees reaching retirement age as specified by their plan become fully vested, regardless of their status in the plan's regular vesting schedule (PBGC 1989, 7).

5. This discussion of pension plan terminations is largely based on PBGC (1989, 18-19) and PBGC (1987, 23-28).

6. Although the number of defined-benefit plans has been declining over time, the number of defined-benefit plan participants has been more stable. Ippolito (1990b) finds that for defined-benefit plans with more than 1,000 participants, which include the vast majority of persons covered by single-employer defined-benefit plans, there has been no shift away from defined-benefit plans. He analyzed trends in defined-benefit and defined-contribution plan coverage over the period from 1979 to 1987.

7. This ambiguity was recognized by Treynor (1977, 633, footnote 3) soon after the enactment of ERISA.

8. Munnell (1982, 27) points out that measuring net assets is problematic for firms that are not publicly traded or are traded infrequently. In these cases net worth must be estimated. Munnell notes that for a firm in Chapter 11 bankruptcy the assumed value of equity in a reorganization plan can be used.

9. The PBGC also has a $100 million line of credit with the U.S. Department of the Treasury. This line of credit has been drawn on only once, for $100,000 in start-up expenses in 1974 (PBGC 1987, 18).

10. The General Accounting Office (GAO) was unable to audit the PBGC's financial statements reported in its *Annual Report* for fiscal years 1990 and 1991 and consequently cautioned "users that the Corporation's [PBGC's] financial

statements have limited reliability." The GAO was unable to evaluate the PBGC's methods for estimating the liability for the present value of future benefits, among other issues (see PBGC 1991c, 53-58; PBGC 1990a, 46-47). PBGC management maintains that it is working to rectify these problems and that its financial statements are accurate (PBGC 1991c, 34-35).

11. These two categories are not included on the balance sheet, but the reasonably probable loss does appear as the most pessimistic case among three scenarios reported in the *Annual Report* since 1989. The annual report gives ten-year forecasts of net losses for the single-employer fund.

12. After terminating its pension plans, LTV negotiated "follow-on" plans, which restored nonguaranteed benefits (the PBGC paying the guaranteed portion), with its union employees and retirees. The PBGC viewed such terminations as an abuse of the pension insurance program (PBGC 1991a).

13. Bodie et al. (1987, 37) point out that the pension benefits of large corporations are not fixed in practice and that these firms appear to share their pension surplus with their retired employees. They believe this practice may reflect a desire to fix the real value of the defined benefit rather than the nominal value.

14. Most terminations before 1974 occurred as a result of corporate reorganizations stemming from bankruptcies and mergers. Also see Bulow (1982, especially 446-47) for a discussion of pre-ERISA pension finance.

15. Marcus (1987) and Estrella and Hirtle (1989) attempt more detailed and realistic modeling of the pension put.

16. $Min(A, B) - B = min(0, A - B) = -max(0, B - A).$

17. Put-call parity is an arbitrage relationship between traded financial claims. The pension put and call as well as the pension liabilities are not traded claims, though in principle markets could develop for such claims.

18. See Ippolito (1985, 1986) for his theory of pension plan underfunding and Ippolito (1988) for his critique of ERISA.

19. In addition to the original 30 percent claim on net worth, SEPPAA increased a plan sponsor's liability to include the difference between 75 percent of unfunded guaranteed benefits and 30 percent of net worth (PBGC 1987, 25). The PPA increased the employer liability to all promised pension benefits. However, this additional claim has general unsecured creditor status in bankruptcy proceedings and may be negotiated with extended payment terms, both of which imply only small additional recoveries to the PBGC. This additional PBGC claim is ignored in the discussion in the text. About 20 percent of the PBGC's claims receive priority treatment in bankruptcy (Lockhart 1991d, 5).

20. Cash accounting is retrospective in recording costs when they are paid, whereas accrual accounting is prospective in measuring costs when they occur. In the case of pension insurance, the accrued cost at a given time is the aggregate value of all pension puts (see Budget 1992, chapter 13).

21. There is some interest in Congress to extend the PBGC's insurance protection to annuities issued by insurance companies. After many standard defined-benefit plan terminations, participants receive annuities, which are not currently insured by the PBGC (see Murphy 1991). Some insurance companies have defaulted on their annuity obligations (most prominently, Executive Life). Insurance companies that are regulated at the state level would have to be regulated at the federal level to avoid moral hazard problems with PBGC coverage of annuities. Such coverage is estimated to increase the PBGC's contingent liability by $50 billion (Lockhart 1991c, 7).

References

Barth, James R., Philip F. Bartholomew, and Michael G. Bradley. "Determinants of Thrift Institution Resolution Costs." *Journal of Finance* 45 (July 1990): 731-54.

Board of Governors of the Federal System. *Balance Sheets for the U.S. Economy, 1945-90*. C.9, Balance Sheets Flow of Funds. Washington, D.C., March 1991.

Bodie, Zvi, Jay O. Light, Randall Mørck, and Robert A. Taggart, Jr. "Funding and Asset Allocation in Corporate Pensions: An Empirical Investigation." In *Issues in Pension Economics*, edited by Zvi Bodie, John B. Shoven, and David A. Wise, 15-44. Chicago: The University of Chicago Press, 1987.

Budget of the United States Government, Fiscal Year 1993. Washington, D.C.: U.S. Government Printing Office, 1992.

Bulow, Jeremy I. "What Are Corporate Pension Liabilities?" *Quarterly Journal of Economics* 97 (August 1982): 435-52.

Estrella, Arturo, and Beverly Hirtle. "The Implicit Liabilities of the Pension Benefit Guaranty Corporation." Federal Reserve Bank of New York Working Paper No. 8905, April 1989.

Garsson, Robert M. "Though Narrower than Expected, Bill Means Lots of Changes." *American Banker*, December 2, 1991, 14.

Harlan, Christi. "LTV, Creditors Link Union Concessions to Fate of Revised Reorganization Plan." *Wall Street Journal*, February 18, 1992, B12.

Haugen, Robert A. "Pension Management in the Context of Corporate Risk Management." *Journal of Portfolio Management* (Fall 1989): 72-78.

Ippolito, Richard A. "The Economic Function of Underfunded Pension Plans." *Journal of Law and Economics* 28 (October 1985): 611-51.

_____. *Pensions, Economics, and Public Policy*. Homewood, Ill.: Dow Jones-Irwin, 1986.

_____. "A Study of the Regulatory Effect of the Employee Retirement Income Security Act." *Journal of Law and Economics* 31 (April 1988): 85-125.

_____. *The Economics of Pension Insurance*. Homewood, Ill.: Irwin, 1989.

_____. "Pension Plan Choice, 1979-1987: Clarifications and Extensions." Pension Benefit Guaranty Corporation. December 1990b.

Kane, Edward J. "Principal-Agent Problems in S&L Salvage." *Journal of Finance* 45 (July 1990): 755-64.

Labaton, Stephen. "Bank Regulator Makes Plea for $70 Billion in U.S. Aid." *New York Times*, November 26, 1991, D1.

Lockhart, James B., III. Testimony before the Subcommittee on Courts and Administrative Practice of the Senate Judiciary Committee. July 24, 1991a.

_____. Testimony before the Task Force on Urgent Fiscal Issues of the House Budget Committee. October 24, 1991b.

_____. Testimony before the Subcommittee on Employment and Housing of the House Committee on Government Operations. October 31, 1991c.

_____. *A PBGC Update*. December 12, 1991d.

LTV Corporation et al. v. Pension Benefit Guaranty Corporation. Cited in LEXIS database as 130 Bankr. 690; 1991 U.S. Dist.; LEXIS 13071, 14 E.B.C. 1225 (Southern District New York, Sept. 13, 1991).

Marcus, Alan J. "Corporate Pension Policy and the Value of PBGC Insurance." In *Issues in Pension Economics*, edited by Zvi Bodie, John B. Shoven, and David A. Wise, 49-76. Chicago: The University of Chicago Press, 1987.

_____, and Israel Shaked. "The Valuation of FDIC Deposit Insurance Using Option-Pricing Estimates." *Journal of Money, Credit, and Banking* 16 (November 1984, part I): 446-60.

Merton, Robert C. "Theory of Rational Option Pricing." *The Bell Journal of Economics* 4 (Spring 1973): 141-83.

_____. "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory." *Journal of Banking and Finance* 1 (June 1977): 3-11.

Munnell, Alicia H. "Guaranteeing Private Pension Benefits: A Potentially Expensive Business." Federal Reserve Bank of Boston *New England Economic Review* (March/April 1982): 24-47.

Murphy, Patricia A. "Congress Takes Aim at Annuity Quality." *Corporate Cashflow*, September 1991, 61-62.

Niehaus, Gregory R. "The PBGC's Flat Fee Schedule, Moral Hazard, and Promised Pension Benefits." *Journal of Banking and Finance* 14 (March 1990): 55-68.

Pension Benefit Guaranty Corporation. *Promises at Risk*. April 1987.

_____. *1988 Annual Report*. 1988.

_____. *Your Pension: Things You Should Know about Your Pension Plan*. Washington, D.C.: U.S. Government Printing Office, 1989.

_____. *1990 Annual Report*. 1990a.

_____. *Study of LBOs and Pension Reversions*. May 1990b.

_____. "PBGC v. LTV." *Background*. October 1991a.

_____. "Top 50 List Shows Increased Underfunding." *News*. November 25, 1991b.

_____. *1991 Annual Report*. 1991c.

_____. "Losses from Underfunded Plans Increased PBGC's Deficit in 1991." *News*. February 18, 1992.

Pesando, James E. "Investment Risk, Bankruptcy Risk, and Pension Reform in Canada." *Journal of Finance* 37 (June 1982): 741-49.

Ronn, Ehud I., and Avinash K. Verma. "Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model." *Journal of Finance* 41 (September 1986): 871-95.

Sharpe, William F. "Corporate Pension Funding Policy." *Journal of Financial Economics* 3 (1976): 183-93.

Treynor, Jack L. "The Principles of Corporate Pension Finance." *Journal of Finance* 32 (May 1977): 627-38.

U.S. Department of Labor. *Pension Protection in Bankruptcy Act of 1991*. Bill submitted to the U.S. Congress. November 21, 1991.

VanDerhei, Jack L. "An Empirical Analysis of Risk-Related Insurance Premiums for the PBGC." *Journal of Risk and Insurance* 57 (June 1990): 240-59.

Westerfield, Randolph, and William Marshall. "Pension Plan Funding Decisions and Corporate Shareholder Value: A New Model and Some Empirical Results." Unpublished Paper. University of Pennsylvania, 1983.

Are There Cost Savings from Bank Mergers?

Aruna Srinivasan

*T*he frequency and scale of U.S. bank mergers are rising. During the past year alone, more than 500 banks announced plans to merge. The total value of transactions completed in 1991 exceeded $20 billion and was nearly five times the $4.3 billion total in 1990. All indications are that this trend will continue and may be a prelude to even more large-scale combinations.

Many policymakers and bankers are viewing mergers as the solution to a number of the problems plaguing U.S. banking: overcapacity, undercapitalization, lack of diversification, and low profitability. Proponents of consolidation argue that during the 1980s banks faced increased competition as deregulation, interest rate volatility, and technological changes in providing financial services diminished the demand for traditional banking products. The nation's banks have had difficulty adjusting to these changes. The industry has failed to contract sufficiently, resulting in growing excess capacity and declining profitability. Consolidation of large banking organizations at the national level is often advocated to reduce excess capacity in banking and to yield cost savings that would enhance banks' profitability.

There is significant evidence that challenges the widely held positive view of consolidation. John H. Boyd and Stanley L. Graham (1991), for example, basing their conclusion on their own analysis as well as other studies, find that big banks in and of themselves are neither more efficient nor safer than moderate-sized banks. Their study concludes that government policies—specifically, the policy of not allowing very large banks to fail—are the driving force behind consolidation. In the same vein, a recent paper by Gary Gorton and Richard Rosen (1991) examines the issue of whether mergers during the 1980s resulted in reduced excess capacity in banking, as measured by total assets. The authors conclude that corporate control problems prevented mergers from serving as an exit mechanism and that

The author is an economist in the financial section of the Atlanta Fed's research department. She thanks Larry Wall for helpful discussions and William Levine for research assistance.

entrenched management leaves failure as the only way assets can be taken from the industry.

The issue of cost savings is important because potential savings are held out to regulators and shareholders alike as justification for recent megamergers. The Federal Reserve System has statutory responsibility to evaluate the likely effects of bank holding company mergers on competition, the financial and managerial resources and future prospects of the firms involved, the convenience and needs of the communities to be served, and Community Reinvestment Act (CRA) requirements.[1] For policy purposes it is important to distinguish between cost savings and improvements in efficiency. Some postmerger cost differences could be simply a function of shrinkage in the resulting firm's size rather than any real improvement in efficiency. However, only those cost savings that result from improved efficiency would create a more competitive institution and constitute a public benefit. Moreover, as Aruna Srinivasan and Larry D. Wall (1992) note, if significant efficiency gains from mergers are available, then blocking mergers because of antitrust concerns, safety and soundness reasons, or CRA considerations could have the negative result of further eroding the competitive position of domestic banks vis-a-vis nonbank firms and foreign banks. If the claimed cost savings can be shown to occur in the typical merger, shareholders should encourage further consolidation and be willing to pay a higher price for a target bank, with the expectation of higher profits from cost savings in the future. On the other hand, if bankers are overestimating the cost savings from mergers, they could risk overpaying for acquisitions because they rely too heavily on cost savings to produce a higher return.

The purpose of this article is to review studies that investigate the question of cost savings associated with bank mergers and to present additional evidence on pre- and postmerger changes in bank efficiency. In particular, it reports findings of recent research suggesting that bank mergers do not necessarily achieve significant economies. This research tests the proposition that bank mergers produced significant savings in cases in which costs are measured as noninterest expenses divided by the sum of net interest revenue and noninterest revenue. The before-and-after performances of all bank mergers completed between 1982 and 1986 are examined, with data analyzed for the full sample and for the fourteen states that make up the Southeast Compact.[2]

The results for the full sample suggest that the median merging banks experienced small but significant decreases in costs during the third and fourth post-merger years. However, when the performance of merged banks is compared with banks that did not merge, the trends in noninterest expenses at sample banks are not found to be significantly different from industrywide trends. Results for the Southeast Compact are similar and indicate decreases in the expense ratio during the second through the fourth postmerger years. As in the full sample, declines in the postmerger expense ratios can be explained by an overall industry decline in expense ratios.

Interesting patterns emerge in the components of noninterest expenses. Both the full sample and Southeast Compact mergers achieved significant reductions in labor costs during the four postmerger years. However, in the case of the full sample, the reduction in salaries was offset by increases in the other-expenses component, suggesting that the banks did not achieve significant economies in consolidating back-office operations. Southeastern banks, on the other hand, were more successful in controlling their postmerger expenses relating to data processing, advertising, product development, sale of branches, and the like.

While the sample of mergers from the 1980s provides the best available data to analyze efficiency gains, two caveats should be added to the conclusions derived in this study. First, the sample is limited to mergers that are much smaller in size than the recently announced megamerger combinations because no comparable mergers occurred during this study's sample period. Second, the acquiring banks' managers may not have intended to reduce costs.

Merger-Related Cost Savings

Estimates. Many parties to the current wave of bank mergers claim that there are substantial cost savings involved. Estimates of cost savings at the level of the industry range from $10 billion to $14 billion. The bulk of these savings ($8 billion) would be achieved through consolidation among the 127 largest bank holding companies (see Bill Atkinson 1991; Simon Brady and Caren Chesler-Marsh 1991).[3] These estimates represent a significant proportion of banks' annual noninterest expenses of $100 billion. Similarly, a recent bank survey concludes that consolidation economies are significant. Thirty-four banks ranging in asset size from $6 billion to more than $40 billion reported saving an average of 35 percent of the target bank's data processing and operations expense within

six to nine months after a merger (Keefe, Bruyette, and Woods Inc. 1990). The study also concluded that it was more difficult to achieve cost savings in a merger of banks of near-equal size. Other estimates by banking industry security analysts suggest that intramarket mergers can reduce expenses of target banks by as much as 40 percent and that intramarket mergers "represent the best hope for sustainable, higher profits in the [banking] industry" (Mark Alpert and Mark Lynch 1991). Savings following out-of-market acquisitions are estimated at 15 percent of the target banks' premerger noninterest expenses.

At the level of individual mergers, the estimates for cost savings range from 23 percent for NationsBank to 32 percent for the proposed merger of BankAmerica and Security Pacific.[4] Some bank analysts believe that bankers are being conservative about potential cost savings and that greater cost savings should eventually translate into higher profits (Gordon Matthews 1991).

Previous Evidence. What does the literature say to expect in terms of cost savings from bank mergers? A number of research papers have investigated issues relating to bank size, consolidation, and efficiency. Most such studies have sought to determine whether larger banking organizations exhibit lower average costs than smaller banks. In general, these studies find significant scale economies at small- to medium-sized banks. The cost advantages for large firms are not as clear. Some studies find that significant scale economies can be realized for banks having as much as $25 billion in assets (William C. Hunter and Stephen G. Timme 1991) while others find diseconomies at institutions with assets beyond about $500 million.[5]

Two recent studies use the cost function methodology to simulate mergers among large banks and estimate the resultant cost savings. Sherrill Shaffer (1991) simulates mergers between banks with more than $1 billion in assets. His results suggest that significant declines in costs occur in only 5 percent of the cases analyzed. Another study of hypothetical mergers among forty-one large banking organizations with substantial branch overlap estimates that even with branch closings the ratio of noninterest expenses to assets would increase (Donald T. Savage 1991). Thus, this line of research has not provided strong evidence suggesting that large mergers in general can be counted on to achieve substantial cost savings.

Another strand of research has attempted to discover whether individual past mergers have resulted in cost savings. While such studies typically focus on the change in noninterest expenses before and after the

merger, changes in profitability and market share are also sometimes examined.[6] Stephen A. Rhoades (1986) examines pre- and postmerger performance using data from the 1970s. The results provide no indication that the performance of the average acquired firm improved after the merger. However, Rhoades's results have limited relevance for recent mergers because during the 1970s deposit-rate ceilings were in effect and banks did not face strong incentives to control noninterest expenses.

Dwight Crane and Jane Linder (1991) estimate changes in noninterest expenses in New England banks during the years from 1982 to 1987. They do not find evidence of substantial cost savings beyond those associated with postmerger shrinkage of the firms in question. Srinivasan and Wall (1992) note a number of limitations with the Crane and Linder study, including the failure to separate intra–holding company mergers from mergers of unaffiliated banks, a restricted sample size, and the failure to include holding company acquisitions.

Srinivasan and Wall examine the changes in noninterest expenses of mergers during the period from 1982 to 1986. The analysis focuses on the merging banks two years before and four years after the merger. The results, which are derived using a combination of univariate and multivariate analysis, indicate that noninterest expenses as a percent of total assets increase after the merger but that the increase can be explained by an overall industry increase in expense ratios and by changes in the composition of the merging banks' balance sheets. There is no evidence that mergers significantly lower expenses.

The univariate analysis compares the ratio of premerger noninterest expenses to total assets with the postmerger ratios. The results suggest that the noninterest-expense ratio increases significantly after the merger. This finding holds regardless of the premerger benchmark used (acquirer financial data or combined acquirer and target institutions' financial data). The median percentage change was found to be highest for large banks (those with more than $10 billion in assets). Midsize banks (between $1 billion and $10 billion in assets) experienced lower-than-average cost increases. The sample was also split in quartiles based on the premerger extent of overlap between acquiring and target banks. The results do not support the hypothesis that intramarket mergers produce significant cost savings. Finally, the data were adjusted for time trends by comparing the pre- and postmerger cost ratios with similar data for banks that did not engage in mergers. After the data were adjusted for industrywide

changes, the merging banks did not display statistically significant increases in their expense ratios.

A regression model is also estimated to control for other potential determinants of noninterest expense such as product mix. The dependent variable in the regression is the dollar value of the combined organization's noninterest expenses during years -2 and -1 and the acquirer's noninterest expenses during years $+1$, $+2$, $+3$, and $+4$. The independent variables include items that generate noninterest expenses such as loans, deposits, and noninterest revenues (used as proxies for fee-based and off-balance-sheet activities). Annual binary variables for the four postmerger years are proxies for the change in costs. In addition, a market overlap variable, time trend variable, and the number of targets the acquirer purchased in year 0 are included.

The regression results do not support the hypothesis that significant cost reductions result from bank mergers. Specifically, the coefficients on the postmerger binary variables are negative and insignificant for the most part. The coefficients on the proxy variables for product-mix and balance-sheet changes are significant with the correct sign. Banks that made multiple acquisitions in the same year were less likely to reduce postmerger costs. While there is some support for the common belief that intramarket mergers produce greater cost savings, the coefficients for the market-overlap variable and the multiple-acquisitions variable offset each other at the mean values of the sample.

Additional Evidence about Cost Savings

This study examines three issues not addressed in Srinivasan and Wall (1992). First, the regression results derived in Srinivasan and Wall's research suggest that product-mix changes are an important determinant of changes in noninterest expenses. However, the univariate results reported in that study do not take into account the effect of product-mix changes. During the 1980s banks increased their reliance on nontraditional, fee-based activities such as data processing, currency trading, and issuance of standby letters of credit, options, and forward contracts. Devoting resources to such activities generates income but does not create financial assets. By contrast, financial intermediation generates both income and financial assets. Consequently, other things being equal, one would expect a bank that is increasing its reliance on off-balance-sheet, fee-based activities simultaneously to increase its ratio of noninterest expense to total assets. In fact, the sam-

ple banks did increase their reliance on nontraditional activities between 1982 and 1986. The rate of growth of noninterest revenue exceeded the rates of growth in total assets and noninterest expenses during the sample period. To allow for the effects of changes in product mix on noninterest expenses this study uses an alternative measure of efficiency, representing the portion of operating income used up in noninterest expenses. Specifically, the ratio equals noninterest expenses divided by the sum of net interest income and noninterest income. Larger numbers reflect inefficiency.

Second, the results are analyzed separately for the full sample and for the fourteen states that make up the Southeast Compact to examine whether southeastern banks displayed different cost-cutting trends following mergers. The formation of the Southeast Compact was based partly on the presumption that interstate mergers would result in cost savings and greater efficiency of the firms involved (see, for example, Larry A. Frieder 1984). The sample period covers the three years preceding and roughly two years following enactment of interstate banking laws in the Southeast.

Third, this study decomposes noninterest expenses into its three components—salaries, premises, and other expenses—to examine whether mergers have a differential impact on the subcategories of expenses.

Data. All bank and bank holding company mergers between 1982 and 1986 were identified using Cates *MergerWatch*; the *Federal Reserve Bulletin*; *Merger Decisions*, published by the Federal Deposit Insurance Corporation; and *Quarterly Reports*, published by the Office of the Comptroller of the Currency (OCC). The sample period was selected to incorporate the effects of deregulation of deposit rates. It was necessary to end the study with mergers completed in 1986 so that there would be four years of data for performance evaluation after the last set of mergers. In order to better capture the effects of larger mergers, the sample includes only those transactions between acquiring and target banks that each exceed $100 million in assets. The data were organized according to the highest-level holding company in a banking organization. Intra–holding company transactions and acquisitions by foreign bank holding companies were excluded from the sample.

Many banks merged often during the sample period, necessitating a number of adjustments to the primary data. First, because year-end financial data were used, multiple acquisitions by a bank holding company during a single year were treated as one transaction. Second, the year the merger actually took place was excluded from the analysis because of insufficient information about whether the purchase method or

pooling-of-interest approach was used to account for the merger.[7] Finally, all transactions in which the acquirer was purchased within two years of the original merger were eliminated from the sample.

The ratio of noninterest expenses to operating income (defined as interest income minus interest expense plus noninterest income) was used to measure efficiency gains (cost savings) from bank mergers. Noninterest expenses were defined as the sum of salaries, premises, and other expenses. Other expenses include net losses from the sale of assets such as branches, the cost of data processing performed for the bank by outside vendors, advertising expenses, directors' fees, deposit insurance premiums, legal fees, and so forth. The financial data were taken from the year-end *Reports of Income and Condition* for the banks and the FR Y-9 Report for the bank holding companies.

To assess the impact of bank mergers, cost ratios were computed for the acquirer and target banks on a pro forma basis, as if they were a single institution during the two years before the merger (years −2 and −1).[8] Another set of cost ratios was computed for the merged bank or bank holding company for the four years after the merger (years +1, +2, +3, and +4).

Table 1 provides summary statistics for the acquirer and target bank holding companies in the full sample as well as the Southeast. Total assets of the southeastern acquiring banks were nearly 55 percent lower than those for the overall sample. The relative size variable (defined as the ratio of the target's assets to the acquirer's assets) was computed to determine whether mergers were among banks of equal size or small banks were being absorbed by larger ones. The mean value of the relative size variable, 0.32 for the full sample (0.30 for the Southeast), suggests that the effects of the merger would not necessarily be hidden among other changes in the acquirer's cost structure.

Table 1
Summary Statistics for Merger Partners
(*Averages in the two years prior to the merger*)

		Mean		Median		Standard Deviation	
	Variable	Full Sample	Southeast	Full Sample	Southeast	Full Sample	Southeast
Total Assets	Acquirer	5,895.06	3,794.54	2,512.17	2,670.81		
(Millions of dollars)	Target	988.21	1,042.67	312.73	388.82		
Noninterest Expenses	Acquirer	167.60	122.32	76.62	90.06		
(Millions of dollars)	Target	32.18	37.44	9.48	12.92		
Operating Income[a]	Acquirer	232.21	176.16	116.93	129.85		
(Millions of dollars)	Target	47.03	49.70	17.59	17.40		
Relative Size[b]		0.32	0.30	0.18	0.17	0.56	0.31
Market Overlap		0.27	0.25	0.09	0.15	0.35	0.30
Noninterest Expense/	Acquirer	0.69	0.69	0.70	0.72	0.07	0.07
Operating Income	Target	0.74	0.76	0.73	0.73	0.16	0.21
Salaries/	Acquirer	0.36	0.36	0.37	0.37	0.05	0.04
Operating Income	Target	0.38	0.39	0.38	0.38	0.08	0.11
Premises/	Acquirer	0.11	0.11	0.11	0.11	0.02	0.02
Operating Income	Target	0.12	0.13	0.12	0.12	0.04	0.04
Other Expense/	Acquirer	0.22	0.21	0.21	0.21	0.04	0.03
Operating Income	Target	0.24	0.25	0.22	0.23	0.08	0.09

[a] Defined as net interest income plus noninterest revenue
[b] The ratio of the target's assets to the acquirer's assets

Bank analysts typically assert that potential cost savings from intramarket mergers far exceed those from out-of-market acquisitions (Alpert and Lynch 1991; Matthews 1992). In this study banking markets were defined as metropolitan statistical areas (MSAs) for urban markets and counties for rural markets. The market overlap variable reflects the extent of premerger deposit overlap between the merger partners and is measured by the share of pro forma deposits derived from overlapping markets. In the full sample the market overlap variable has a mean of 0.27, suggesting that, on average, 27 percent of the combined organization's deposits were derived from markets in which both the acquirer and target banks were operating during the year before the merger.

The ratio of noninterest expense to the sum of net interest income and noninterest revenue indicates how efficiently banks spend money to make money. For example, the average acquiring bank spent $.69 cents to produce $1.00 of total net interest and noninterest revenue. According to this measure, target banks are not as efficient as acquiring banks, on average. Salaries are the most important component of noninterest expenses, followed by other expenses and premises. Any economies of scale in marketing and back-office operations would show up as declines in noninterest expenses and its components. Hereafter, the ratio of noninterest expense to the sum of net interest income and noninterest income is referred to as the noninterest expense ratio, or simply expense ratio.

Results. Table 2 presents the mean percentage change in the noninterest expense ratio and its components for the full sample and the Southeast. The results for the full sample suggest that the merging banks

Table 2
Mean Percentage Change in Noninterest Expense Ratio
(and Its Components) of Merged Banks
(t-ratios in parentheses)

Expense Ratio	Year +1 Change	Year +2 Change	Year +3 Change	Year +4 Change
Noninterest Expense				
Full Sample	0.04	−0.75	−1.60	−1.14
	(0.06)	(−0.88)	(−1.53)	(−0.83)
Southeast	−1.30	−3.18	−3.49	−4.39
	(−1.08)	(−2.48)*	(−2.05)*	(−2.98)**
Salaries				
Full Sample	−2.49	−4.41	−6.52	−7.01
	(−3.78)**	(−5.86)**	(−7.92)**	(−7.36)**
Southeast	−4.19	−5.92	−6.73	−7.37
	(−5.15)**	(−5.54)**	(−5.13)**	(−5.80)**
Premises				
Full Sample	0.48	0.11	−0.48	0.45
	(0.26)	(0.05)	(−0.22)	(0.18)
Southeast	−5.93	−4.59	−4.70	−4.85
	(−2.57)*	(−1.95)	(−1.82)	(−1.64)
Other Expenses				
Full Sample	6.06	7.23	8.02	9.27
	(3.95)**	(3.98)**	(4.13)**	(4.54)**
Southeast	6.45	3.08	4.42	3.00
	(2.10)*	(1.04)	(1.21)	(1.41)

Note: *The denominator is defined as interest income minus interest expense plus noninterest income. The premerger benchmark is the combined noninterest expenses for the merger partners averaged over years −2 and −1. Number of observations: full sample = 240; Southeast = 77.*

* *Indicates significance at the 5 percent level*
** *Indicates significance at the 1 percent level*

experienced small, insignificant declines in their noninterest expense ratio during years +2 through +4. On the other hand, banks in the Southeast managed to reduce their noninterest expense ratio by nearly 5 percent relative to the premerger period. Virtually all of the decline in the noninterest expense ratio can be attributed to declines in the salaries component, which was statistically significant for both the full sample and the Southeast. The ratio of premises expenses to operating income showed declines in the Southeast but was not statistically significant. Finally, other expenses, the second-largest component of noninterest expenses, showed significant increases growing from 6.06 percent (year +1) to 9.27 percent (year +4) in the full sample. Merging banks in the Southeast experienced statistically insignificant increases in the other-expenses component. Splitting the sample by asset size class reveals that large banks (exceeding $10 billion in assets) displayed statistically insignificant increases in expense ratios while medium-sized banks ($1 billion to $10 billion in assets) experienced significant reductions in expense ratios during the first four postmerger years.

From a statistical point of view, it is important to compute changes using the sample median in addition to the mean because mean values could be influenced by unusually high or low values.[9] Furthermore, testing the significance of means requires assumptions about the distribution of the underlying population—for example, that the population is normally distributed. This assumption may not be valid in the case of small samples.

Table 3 reports median percentage changes in the noninterest expense ratio and its components for the first four postmerger years. The Wilcoxon signed rank statistic was used to test for the significance of the percentage changes.[10] Declines in the noninterest expenses and salaries ratios are greater in magnitude than those reported in Table 2 and are statistically significant for the most part. Increases in other expenses are statistically significant but are of smaller magnitude than those reported in Table 2.

The results presented in Tables 2 and 3 suggest that the merging banks were successful in reducing the salaries component of noninterest expenses. However, that reduction was at least partially offset by rapid increases in the other-expenses component, resulting in unchanged or slightly lower noninterest expenses for the overall sample. Southeastern banks, on the

---

Table 3
Median Percentage Change in Noninterest Expense Ratio
(and Its Components) of Merged Banks

Expense Ratio	Year +1 Change[a]	Year +2 Change[a]	Year +3 Change[a]	Year +4 Change[a]
Noninterest Expense				
Full Sample	−0.30	−1.87	−3.96*	−4.84*
Southeast	−1.14	−3.40**	−5.12**	−5.18**
Salaries				
Full Sample	−3.73**	−5.83**	−7.99**	−9.32**
Southeast	−4.26**	−5.98**	−7.36**	−9.18**
Premises				
Full Sample	−1.48	−4.25	−5.31	−5.32
Southeast	−4.31*	−5.93	−6.71	−7.81
Other Expenses				
Full Sample	4.60**	3.32**	3.04**	3.48*
Southeast	2.96	0.77	−0.28	1.01

Note: *The denominator is defined as interest income minus interest expense plus noninterest income. The premerger benchmark is the combined noninterest expenses for the merger partners averaged over years −2 and −1.*

a *Significance levels are based on two-tailed Wilcoxon signed rank tests.*

* *Indicates significance at the 5 percent level*

** *Indicates significance at the 1 percent level*

other hand, controlled growth in the other-expenses component and, consequently, experienced statistically significant reductions in noninterest expenses. The results imply that, while banks in the full sample eliminated redundant management positions and reduced staff, they realized limited if any economies from consolidating information systems (computer hardware, software, and facilities) and in marketing activities such as product development and advertising.

A potential problem for evaluating the pre- and postmerger changes is that some of the differences could be accounted for by industrywide factors. One way to isolate merger-specific changes is to compare the sample of merging banks with a relevant industry group. In this case the relevant industry group includes all banks that did not merge. The industry-adjusted data for the full sample are reported in Table 4.[11] The first column in Table 4 reports the premerger noninterest expenses ratio (and its components), averaged over

years -2 and -1. The t statistic tests the null hypothesis that the sample and industry banks are not significantly different before merger. For the most part, there are no significant differences in the premerger expense ratios across the sample and industry in years -2 and -1. The only exception is the average for other expenses, which is significantly lower at the merging banks. Columns 2 through 5 report the postmerger changes in noninterest expenses and its components. The t statistic tests the hypothesis that the postmerger changes in noninterest expenses at the sample banks are not significantly different from industrywide trends in noninterest expenses during the same period. The null hypothesis cannot be rejected for any of the postmerger years, suggesting that the merging banks were no more efficient than the industry.[12]

The sample banks achieved significantly larger salary reductions than the comparison group. Changes

Table 4
Performance of Merged Banks Compared with the Industry

Expense Ratio[a]	Averages over Years -2 and -1	Year +1 Change[b]	Year +2 Change[b]	Year +3 Change[b]	Year +4 Change[b]
Noninterest Expenses					
Sample	0.70	0.0004	−0.0075	−0.0160	−0.0114
Industry	0.71	0.0036	0.0022	0.0007	0.0080
t-Test of Mean Differences	1.36	0.47	1.03	1.47	1.25
Salaries					
Sample	0.37	−0.0249	−0.0441	−0.0652	−0.0701
Industry	0.36	−0.0189	−0.0259	−0.0317	−0.0297
t-Test of Mean Differences	−1.82	0.81	2.16*	3.62**	3.81**
Premises					
Sample	0.11	0.0048	0.0011	0.0048	0.0045
Industry	0.11	0.0421	0.0493	0.0423	0.0424
t-Test of Mean Differences	0.64	1.82	2.14*	1.94	1.41
Other Expenses					
Sample	0.22	0.0606	0.0723	0.0802	0.0947
Industry	0.23	0.0512	0.0607	0.0710	0.0940
t-Test of Mean Differences	4.19**	−0.53	−0.56	−0.43	0.05

[a] The denominator is defined as net interest income plus noninterest income.
[b] These numbers represent actual changes in noninterest expenses and its components (not percentage changes).
* Indicates significance at the 5 percent level
** Indicates significance at the 1 percent level

in premises and other expenses at sample banks did not differ significantly from the industry. Overall the data suggest that merging banks did not improve their noninterest expenses relative to the industry.

The results of testing the view that intramarket mergers can result in significant cost savings are presented in Table 5.[13] The sample is split by quartile of the market overlap variable.[14] The first quartile includes merger observations in which there was no market overlap. The second quartile includes observations with overlap greater than zero but less than the median and so on for the third and fourth quartiles. A value of 100 indicates perfect overlap. In general, the results for the full sample do not support the claim that a higher degree of market overlap is associated with greater cost savings. Southeastern banks in the first

and third quartiles displayed significant reductions in noninterest expenses.

There are several possible explanations for this result. First, market overlap as measured here may overestimate the actual branch overlap. Savage (1991) estimates that only 2.7 percent of U.S. bank deposits are in "duplicative" branches of large banks.[15] Second, branch operating costs may be low, and any efficiency gains from closing overlapping branches may be small in relation to the overall banking system. Third, controlling for other potential determinants of noninterest expenses, Srinivasan and Wall (1992) find that opportunities for cost-cutting are greater when the merger partners operate in the same deposit market. This relationship may not be apparent in the univariate analysis because the market overlap variable

Table 5
Mean Percentage Change in the Noninterest Expense Ratio
of Merged Banks by Market Overlap Quartile
(t-ratios in parentheses)

Market Overlap Quartile	Year +1 Change	Year +2 Change	Year +3 Change	Year +4 Change
First Quartile				
Full Sample	0.39	−0.39	−2.65	−0.30
(No Overlap)	(0.46)	(−0.23)	(−1.99)	(−0.15)
Southeast	−2.57	−5.37	−6.03	−4.73
(No Overlap)	(−2.09)*	(−4.11)**	(−3.53)**	(−2.33)*
Second Quartile				
Full Sample	−0.74	−1.96	−3.84	3.70
(0−8.83)	(−0.63)	(−1.50)	(−2.43)**	(0.87)
Southeast	2.20	0.35	0.97	−4.15
(0−15.12)	(0.48)	(0.08)	(0.15)	(−1.22)
Third Quartile				
Full Sample	−0.02	−1.21	−0.57	−2.94
(8.83−53.37)	(−0.01)	(−0.85)	(−0.25)	(−1.37)
Southeast	−2.95	−5.38	−5.55	−7.03
(15.12−37.24)	(−1.36)	(−2.54)**	(−2.44)**	(−3.06)**
Fourth Quartile				
Full Sample	0.02	−0.01	0.00	−3.18
(53.37−100)	(0.01)	(−0.06)	(0.00)	(−0.98)
Southeast	−0.94	−1.19	−2.11	−1.35
(37.24−100)	(−0.49)	(−0.41)	(−0.62)	(−0.33)

Note: *The denominator is defined as net interest income plus noninterest income. The premerger benchmark is the combined noninterest expenses for the merger partners averaged over years −2 and −1.*

 * Indicates significance at the 1 percent level
** Indicates significance at the 1 percent level

is correlated with other variables such as noninterest revenue or total assets.

Bank surveys typically disclose that in a merger of equals neither bank has the capacity to absorb the data-processing and back-office operations of the other. In addition, political considerations often make prompt action difficult (Keefe, Bruyette, and Woods Inc. 1990; Brady and Chesler-Marsh 1991).[16] Table 6 splits the sample by relative size of the merger partners to examine how successful mergers of equals have been in reducing costs. Relative size is measured by the ratio of the target's premerger assets to the acquirer's premerger assets. Thus, the first quartile in Table 6 includes merger observations in which the target's assets amounted to less than 9 percent of the acquirer's, whereas the fourth quartile includes transactions in which the target's assets exceeded 40 percent of the

acquirer's premerger assets. Contrary to popular notion, there are significant reductions in costs in the third and fourth quartiles, both for the full sample and for institutions in the Southeast Compact. One possible explanation is that mergers of equals generally involved small and medium-sized banks (less than $5 billion in assets) during the sample period. The economies of scale literature has consistently found significant cost savings opportunities for banks in this size category.

Conclusion

This study examines the noninterest expenses of bank merger partners for the two years prior to the

Table 6
Mean Percentage Change in the Noninterest Expense Ratio
of Merged Banks by Relative Size Quartile
(*t-ratios in parentheses*)

Relative Size Quartile	Year +1 Change	Year +2 Change	Year +3 Change	Year +4 Change
First Quartile				
Full Sample	1.23	0.55	2.25	5.25
(0−8.12)	(1.16)	(0.38)	(0.88)	(1.53)
Southeast	−2.08	−2.98	0.31	−2.39
(0−7.71)	(−1.42)	(−1.01)	(0.06)	(−0.81)
Second Quartile				
Full Sample	0.49	−0.33	0.70	−1.02
(8.12−20.27)	(0.53)	(−0.26)	(0.31)	(−0.41)
Southeast	0.33	0.25	−1.01	−1.00
(7.71−19.23)	(0.21)	(0.10)	(−0.39)	(−0.27)
Third Quartile				
Full Sample	−0.31	0.22	−3.20	−1.68
(20.27−43.40)	(−0.21)	(0.10)	(−2.43)**	(−0.67)
Southeast	−0.77	−4.50	−6.33	−8.05
(19.23−39.40)	(−0.19)	(−1.60)	(−2.41)**	(−3.26)**
Fourth Quartile				
Full Sample	0.17	−2.20	−4.95	−6.30
(43.45−1.50)	(0.12)	(−1.23)	(−2.46)**	(−3.26)**
Southeast	−1.79	−4.61	−5.65	−5.26
(39.40−1.22)	(−1.10)	(−2.07)*	(−2.79)**	(−2.03)*

Note: *Relative size = target's premerger assets/acquirer's premerger assets. The denominator is defined as net interest income plus noninterest income.*
* *Indicates significance at the 5 percent level*
** *Indicates significance at the 1 percent level*

merger and four years after the merger. The sample consists of all mergers between 1982 and 1986 involving participants that had total assets of at least $100 million. Efficiency gains (cost savings) are measured relative to the sum of net interest income and noninterest revenue. Merger data for Southeast banking organizations are analyzed to determine whether relaxation of interstate banking laws in this region contributed to significant efficiency gains at merging banks.

The results generated by comparing the pre- and postmerger expense ratios suggest that the average bank merger produced small but significant declines in the ratio of noninterest expense to the sum of net interest income and noninterest income. For example, the median merging bank in the full sample reduced costs by 4.84 percent (in the Southeast, by 5.18 percent) by the fourth postmerger year. The decline in expenses for the average (mean) merging bank in the full sample was statistically insignificant. The magnitude of decline in noninterest expenses four years after the merger was well below the projected declines of 23 percent to 32 percent for recent megamergers. The components of noninterest expenses display interesting patterns and suggest that, while acquiring banks in the full sample and the Southeast have been successful in reducing the salaries component, these savings have been largely offset for the full sample by increases in the other-expenses component; southeastern banks, on the other hand, controlled growth in this component.

The data were adjusted for time trends using an industry comparison group. Changes in noninterest expenses at the merging banks were not significantly different from industrywide trends during the same period. The univariate analysis did not find substantial cost reductions in past intramarket mergers, although the effect may be masked because of correlation between the market overlap variable and other potential determinants of noninterest expenses. Mergers of equal size were found to show greater-than-average reductions in costs.

The analysis presented in this paper suggests the need to control for changes in product mix when measuring operational efficiency. The univariate results derived in this study allow for differences in product mix at the sample banks. The choice of denominator (total assets or net interest income plus noninterest revenue) affects the pre- and postmerger univariate results. At the same time, the results are consistent with those derived in Srinivasan and Wall (1992) because adjusting for industrywide trends reveals that the typical merging bank performs no better than the industry. Further, the regression analysis in Srinivasan and Wall controls for changes in product mix over time and concludes that the typical bank merger does not reduce operating costs significantly.

An analysis of past bank mergers indicates that significant cost savings are far from automatic. Researchers, bankers, and bank analysts all agree that regardless of size there are wide variations in banks' efficiency, so the potential for cost savings may be greater in some mergers than in others. While this study does not conclude that cost savings are impossible to achieve, its findings suggest that regulators and shareholders should give even more careful consideration to claims of cost savings on a case-by-case basis.

Notes

1. The Federal Deposit Insurance Corporation and the Office of the Comptroller of the Currency also observe similar standards in evaluating bank mergers.
2. The Southeast Compact includes the following states: Alabama, Arkansas, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia. See Goudreau and Wall (1990) for a description of interstate banking trends in the Southeast.
3. Of the $10 billion in cost savings mentioned in the Treasury proposal, $8 billion would be achieved through consolidation among the 127 largest bank holding companies in the United States, and $2 billion would be saved through repeal of the McFadden Act. McKinsey and Company adds $4 bil-

lion in savings from small bank consolidations to the Treasury estimate.
4. Srinivasan and Wall (1992) derive these estimates by taking management's projected savings and dividing by Value-Line's projection of the target banks' noninterest expenses for 1991.
5. A recent review of the evidence is in Humphrey (1990).
6. Spindt and Tarhan (1991) examine the issue of synergistic gains from bank mergers. Using data on 300 mergers that occurred in 1986, they find that prior to mergers target banks tended to be underperformers, and acquiring banks generally were normal performers. The combined institution tended to increase its return on equity after the merger. Spindt and Tarhan interpret their results as being consistent

with an efficiently operating market for corporate control. Their results have limited relevance for current merger trends because the median independent target bank in their sample had only $26 million in assets.

7. Some mergers are accounted for using the purchase method, by which accounts are consolidated after the merger, while others use a pooling-of-interest approach that consolidates accounts from the beginning of the year. Results for year 0 are therefore not comparable across banks.

8. This benchmark indicates the extent to which the acquirer transferred its management style and efficiency skills to the target bank.

9. While the mean is the average value of the sample, the median is the value such that half of the numbers in a list are above it and half are below it.

10. The Wilcoxon signed rank statistic was constructed in the following manner (Hollander and Wolfe 1973). First, the absolute values of the percentage changes were ranked. Then the sums of the ranks were calculated for the positive differences and negative differences, giving T_+ and T_-. If T is the smaller of T_+ and T_- and the sample size N is large, then T is a normal random variable with parameters: mean $\mu_T = [N(N + 1)]/4$ and standard deviation $\sigma_T = [N(N + 1)(2N + 1)]/24$. The statistic $(T - \mu_T)/\sigma_T$ is a standard normal variable that can be compared with a standard normal table to decide whether to accept the null hypothesis, that the percentage changes are not significant.

11. The analysis was also carried out for merging banks in the Southeast. The relevant comparison group included banks in the Southeast Compact that did not engage in mergers. The results are consistent with those derived using the full sample.

12. Although the changes in the expense ratios for the sample and industry groups have opposite signs during years +2, +3, and +4, there was a great deal of variability within the industry, so the claim that the industry and sample bank expense changes are identical could not be rejected.

13. The results using medians were very similar and are not reported here.

14. There are unequal numbers of observations in the quartiles because the controlling variable used to create the quartiles is market overlap.

15. Savage (1991) measures branch overlap within five-digit postal zip code areas.

16. The sheer size of the mergers creates concerns about employee morale, turnover, and balance of power. Many of these issues paralyze banks and prevent them from taking prompt, decisive actions.

References

Alpert, Mark, and Mark Lynch. "Bank Shots: Intramarket Mergers, Part II." Bear Stearns and Co., *Research Highlights*, December 13, 1991.

Atkinson, Bill. "Cost Reductions under Reform Widely Debated." *American Banker*, April 19, 1991, 1, 8.

Boyd, John H., and Stanley L. Graham. "Investigating the Bank Consolidation Trend." Federal Reserve Bank of Minneapolis *Quarterly Review* 15 (Spring 1991): 3-15.

Brady, Simon, and Caren Chesler-Marsh. "The Madness of Mergers." *Euromoney*, September 1991, 66-80.

Crane, Dwight, and Jane Linder. "Merging versus Branching: the Pros and Cons." *American Banker*, April 17, 1991, 4.

Frieder, Larry A. "The Southern Financial Marketplace: The Florida Interstate Banking Study." In *Proceedings of a Conference on Bank Structure and Competition, April 23-25, 1984*, 12-34. Chicago: Federal Reserve Bank of Chicago, 1984.

Gorton, Gary, and Richard Rosen. "Overcapacity and Exit from Banking." The Wharton School. Mimeo, November 1991.

Goudreau, Robert G., and Larry D. Wall. "Southeastern Interstate Banking and Consolidation: 1984-89." Federal Reserve Bank of Atlanta *Economic Review* 75 (November/December 1990): 32-54.

Hollander, Myles, and Douglas A. Wolfe. *Nonparametric Statistical Methods*. New York: John Wiley and Sons, 1973.

Humphrey, David B. "Why Do Estimates of Bank Scale Economies Differ?" Federal Reserve Bank of Richmond *Economic Review* 76 (September/October 1990): 38-50.

Hunter, William C., and Stephen G. Timme. "Technical Change in Large U.S. Commercial Banks." *Journal of Business* 64 (April 1991): 206-45.

Keefe, Bruyette, and Woods Inc. "The Economic Effects of Mergers and Acquisitions on Consolidating Data and Operations Centers." A survey conducted by Keefe, Bruyette, and Woods Inc., 1990.

Matthews, Gordon. "Why NCNB Paid a Rich Price for C&S." *American Banker*, July 24, 1991, 16.

_____. "Merger Boom of 1991 Will Test Wisdom of In-Market Mergers." *American Banker*, January 14, 1992, 1, 6, 8.

Rhoades, Stephen A. "The Operating Performance of Acquired Firms in Banking before and after Acquisition." Board of Governors of the Federal Reserve System, Staff Study 149, April 1986.

Savage, Donald T. "Mergers, Branch Closings, and Cost Savings." Board of Governors of the Federal Reserve System. Mimeo, May 1991.

Shaffer, Sherrill. "Potential Merger Synergies among Large Commercial Banks." Federal Reserve Bank of Philadelphia Working Paper 91-17, October 1991.

Spindt, Paul A., and Vefa Tarhan. "Are There Synergies in Bank Mergers?" Paper presented at the annual meeting of the Southern Finance Association, Key West, Florida, November 20-23, 1991.

Srinivasan, Aruna, and Larry D. Wall. "Cost Savings Associated with Bank Mergers." Federal Reserve Bank of Atlanta Working Paper 92-2, February 1992.

FYI
Path-Dependent Options

William C. Hunter and David W. Stowe

A relatively new class of options—the so-called path-dependent options—has become increasingly popular in recent years. Like other options, these contracts give their owners the right—but not the obligation—to buy or sell a specific quantity of an underlying asset (stock, bond, futures contract, commodity, and so forth) at a specified price, called the strike or exercise price, during a specific time period.

Since 1982 the use of path-dependent options has grown dramatically. A path-dependent option has a payout directly related to movements in the price of the underlying asset during the option's life. In principle, these options take many forms and can be contingent on virtually any statistic of the underlying asset's price path—for example, the high price, the low price, or the average price over some time period. Today path-dependent options are available on a host of assets including common stock, interest rate products, precious metals, commodities, foreign currencies, and stock indexes; they are often used with convertible securities issues and in merger transactions and have recently begun trading on two major exchanges.[1]

In many cases these options allow investors to limit their potential losses (and gains) and thus have a type of built-in insurance feature. They also allow investors with specialized knowledge about asset price volatility to exploit this information better in their investing and hedging activities. While factors such as cost or risk mean that path-dependent options will not satisfy every investor's needs, these options have generated interest by filling several voids or niches in derivative securities markets.

The sections that follow introduce the notion of path dependency, review the modern origins of path-dependent options, and give several examples of reasons that investors and institutions find these options attractive. The discussion also describes some essential features of three types of path-dependent

Hunter is vice president and senior economist in charge of basic research in the Atlanta Fed's research department. Stowe is a financial analyst in the department's macropolicy section.

Federal Reserve Bank of Atlanta

Economic Review   **29**

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

options—the lookback option, the barrier option, and the average-rate or Asian option. A forthcoming article in this *Economic Review* will describe in detail the valuation and pricing of these options, illustrate how they are used by individual investors and firms, and discuss their advantages and disadvantages (risks) as investment vehicles.

The American Put Option

A call option conveys to its owner the right to buy the underlying asset while a put conveys the right to sell the underlying asset. An option that allows its owner to buy or sell the underlying asset (exercise the option) at any time during the life of the option is called an American option. An option allowing the owner to exercise his or her right only at the option's expiration or maturity date is called a European option.

The payoff on a European call or put option written on a share of common stock that pays no dividends depends only on the market price of the underlying common stock at maturity. That is, if T is the maturity date of the option, t is the current date, X is the exercise price, and S is the market price of the underlying common stock at time $T > t$, then the payoff or intrinsic value at T is equal to the larger of the quantities $(S - X)$ and zero for the call option and $(X - S)$ and zero for the put option. Using standard options notation, the payoff on the European call at time T is written as $max(S - X, 0)$ while the payoff on the European put at time T is equal to $max(X - S, 0)$. Payoff at time T, $(T > t)$, on the European call or put option on a stock that pays no dividends is independent of the particular path taken by the stock price during the period between the times t and T. Such standard European call and put options on a nondividend paying stock are the simplest examples of what are called path-independent options.

In contrast, an American put option written on a share of common stock has a path-dependent payoff structure. (This happens to be the case irrespective of whether the stock pays a dividend or not.) For example, looking forward from the perspective of date t, the payoff at time $T > t$ on the American put option depends not only on the price of the underlying stock at time T but also on the particular time path followed by the stock between times t and T.

An illustration will demonstrate the straightforward intuition behind the statement that the American put is an example of a path-dependent option. Assume that the underlying stock pays no dividend and that the put option is in the money—that is, the market price of the stock is less than the exercise price. The investor holding the put could exercise the option and receive an amount of cash equal to the exercise price minus the current price of the stock that he has just sold, $X - S$. In turn, this cash can be invested at the risk-free rate of interest to earn money during the remaining life of the option. At expiration the investor receives the amount $X - S$, his original investment, plus the interest earned over the remaining life of the option. An investor choosing not to exercise the put and waiting until expiration would receive only the amount $X - S$. It should be obvious that if the stock price is close enough to zero at the date the investor chooses to exercise early, he or she will be better off; the principal and interest received from investing the proceeds will exceed the difference between the exercise price and the stock price at the option's maturity date. In addition, the cases in which early exercise is optimal occur when the put is selling for $X - S$ so that selling it would be less profitable than exercising it and investing the proceeds. The key condition making early exercise preferable is that the stock price follows a path that drops close enough to zero over the life of the option to make the principal and interest earned by exercising the option greater than the exercise price minus the stock price at maturity. Thus, the payoff to the investor is seen as path dependent.

The Modern Origins of Path-Dependent Options: The Lookback Option

For both standard options and securities, specific examples having characteristics similar to path-dependent options can probably be traced back at least to the early 1800s. However, the modern treatment of these securities—the rigorous valuation or pricing of these claims on the basis of dynamic hedging principles—is a more recent phenomenon, set in motion in 1979 with publication of an article by M. Barry Goldman, Howard Sosin, and Mary Ann Gatto. The authors had derived an explicit valuation formula for a hypothetical option epitomizing the age-old finance dictum of buy low (cheap) and sell high (dear)—the so-called lookback option. To allow buying low and selling high, the exercise price on the lookback option is set at the expiration of the contract instead of at contract origination (as it is for standard options). That is, at expiration the owner could "look back over the life of the option"

and choose as the exercise price the most favorable price that had occurred.

If a lookback call option were exercised, the owner would be able to buy the underlying asset at the lowest price that occurred during the life of the option. Similarly, the owner of a lookback put would be able to sell the underlying asset at its highest price realized over the life of the option. It is clear that the payoff on a lookback option depends not only on the underlying asset's price on the expiration date of the option but also on the particular path followed by the price of the asset over the life of the option, hence the path dependency.

Countering the argument that their research was a purely hypothetical exercise in contingent claims valuation, Goldman, Sosin, and Gatto (1979) argued that lookback options could be of value to investors as speculative and hedging instruments and could survive as traded securities. In less than two-and-a-half years the authors were proven correct. On March 16, 1982, Macotta Metals Corporation of New York introduced and began trading lookback options on gold, silver, and platinum. The lookback call gave an investor the right to buy gold, silver, or platinum at its ex post realized low price, and the lookback put allowed the investor to sell the precious metal at its ex post realized high price.

Uses of Path-Dependent Options

The choice of the particular price-path statistic on which a path-dependent option is based depends on the motivation of the option writer, ranging from wanting to control some particular risk to filling some niche in the market. Some specific examples illustrate this point.

On April 22, 1982, Manufacturers Hanover Corporation sold a $100 million note offering. The sale required holders to convert the securities at maturity in 1992 into shares of the company's common stock. The conversion price would be the lower of $55.55 and the average closing price of the common stock for the thirty-day period immediately preceding the notes' maturity. By making the conversion price dependent on the average price of the common stock, the company alleviated suspicions among investors that management would fraudulently manipulate the stock price upward just before the conversion date.[2]

The "capped" stock-index option is an example of an exchange-traded path-dependent option developed to fill a special niche or appeal to specific investors in the market. Capped stock-index options, fairly new examples of path-dependent options, are so named because they place ceilings on profitability. Because of these ceilings, capped options are cheaper than traditional stock-index options. Capped options were launched during the fall of 1991 on both the Chicago Board Options Exchange and the American Stock Exchange. Like other index options, they can be used to protect the values of stock portfolios by providing a cheaper way to obtain portfolio insurance.

Capped options trade off the Standard and Poor's (S&P) 100 and 500 Indexes on the CBOE and the Major Market and Institutional Indexes on the Amex. Like the standard call option, the value of a capped call option increases if its underlying index goes up, and a capped put option's value increases if the index declines. If the underlying index fails to attain the level specified by the option contract, known as the strike price, the options expire worthless and the sellers keep all of the premiums they collected. On the other hand, if the indexes reach the strike price, sellers must pay the optionholders the difference between the index level and the strike price but no more than a fixed cap value.

Each of these options has a cap price. For the options currently trading on the CBOE the cap price is set thirty points above the strike price for a call and thirty points below the strike price for a put, giving the options a cap value of $3,000 (thirty points times $100 per point). For those trading on the Amex the cap price is set at twenty points above and below the strike price, yielding a cap value of $2,000 (20 points times $100 per point). The purpose of the cap price is to force automatic exercise of the options. If the underlying index closes at or above the cap price for a call option or at or below the cap price for the put option, the options are automatically exercised and the cap buyers are paid the cap value two days after exercise.

The following scenario illustrates the mechanics of the capped option. An investor believes that the stock market will rally modestly from its closing value of 378 for the S&P 500 index on, say, January 12, 1992, and the third Friday in March 1992, the expiration date for the cap. The strike or exercise price on capped calls is 390, making the cap price 420 (390 plus the thirty points for the S&P 500 index). If the index closes at or above 420 between January 12 and the third Friday in March 1992, the capped call purchaser will be paid $3,000 (the net profit would be less by the amount of the premium). If the index closes at a figure less than 420 but greater than the strike price of 390—for example, 400—the purchaser will be paid an

amount equal to the value of the index minus the strike price, in this case ten points times $100 per point or $1,000. On the other hand, if the index fails to reach the strike price of 390, the option expires worthless and the seller keeps the entire premium collected.

As is true for other exchange-traded options, the owners of capped options can sell them in the open market before maturity. Clearly, the payoff on the capped option depends on the particular path the underlying index follows over the life of the option; the option is path dependent. One appealing characteristic of the capped put or call option is that the seller's or writer's risk is limited to the cap amount or value while theoretically there is no limit to the risk faced by the writer or seller of a standard stock-index or -equity option. This feature of capped options, which is essentially a kind of built-in insurance, should make investors more willing to write options on the indexes offering them.

Other Popular Path-Dependent Options

The average-rate option and the barrier option are two other frequently used path-dependent options that are growing in popularity. Both are currently used most extensively in the foreign exchange markets. However, their structure is such that their use will most likely increase in domestic markets in the future. The capped option discussed above exhibits some of the essential features of a barrier option.

Barrier Options. Simply stated, a barrier option is a path-dependent option that is either canceled, activated, or exercised if the underlying instrument (the stock index in the case of a capped index option) reaches a certain level, regardless of the point at which the underlying asset is trading at maturity. Barrier options, also known as knock-out, knock-in, or trigger options, are typically straight European options until or from the time the underlying instrument reaches the barrier price.

There are four popular types of barrier options: up-and-out, up-and-in, down-and-out, and down-and-in. With the up-and-out barrier, the option is canceled should the underlying instrument rise above a certain level. The up-and-in option, on the other hand, is worthless unless the underlying instrument rises above a certain level or price, at which point it becomes a normal put option. Down-and-out options are canceled if the underlying instrument falls below a certain price. Down-and-ins are activated only when the underlying instrument's price falls to a certain level.

Because of these extinguishing or activating features, barrier options are cheaper than ordinary European options and are thus attractive to investors who are averse to paying large premiums. In addition, as illustrated in the case of the capped option, the sellers or writers of barrier options may be able to limit their downside risk.

Average-Rate Options. Average-rate or Asian options are path-dependent options, European in structure, for which the strike price is based on the average (geometric or arithmetic) price of the underlying instrument over a specified period of time, so the actual strike price is not determined until the exercise date on the contract. For foreign exchange average-rate options, the actual practice is for the average to be taken from the option's start date to a preagreed setting date. For example, suppose that a U.S. exporter buys an average-rate floating-strike call option to purchase a foreign currency for U.S. dollars at the average exchange rate over some given period, with the option expiring at the end of the period. If the average exchange rate over the period is less than the spot exchange rate at the time payment is due to the foreign importer, the exporter would profit more from exercising the option than transacting at the spot exchange rate. On the other hand, if the period's average exchange rate exceeds the spot rate, the exporter is better off converting dollars at the current spot exchange rate, in which case the option expires worthless. This example also shows that it is possible to use average rate options to hedge or limit the uncertainty associated with regular foreign cash inflows and outflows as a result of volatile exchange rates.

Many multinational corporations use average-rate put options on foreign currencies to hedge their estimated monthly foreign exchange income in an effort to achieve some budgeted average exchange rate for the year. Hence, the design of this particular option is of great value to these corporations that are in the market on a regular basis. Current accounting principles provide for foreign currency transactions to be translated at either the spot rate at the time of the transaction or the spot rate for the date of the firm's balance sheet. Any variations can be flowed through into the firm's income. For a path-dependent put, the option can be exercised if the balance-sheet rate is less than the strike (average) rate, resulting in the appearance of additional income. This additional income is calculated by multiplying the nominal amount by the difference between the strike rate and the spot rate and subtracting from this figure the amount of premium paid.

Large multinational commercial banks offer average-rate currency options to their multinational customers because these companies' usual spot dealings leave them with an average exchange rate on their books. By selling path-dependent average-rate currency options, the banks offset the average-rate foreign exchange risk exposure on their books. The premiums banks receive enhance yield by reducing their funding costs or by lowering their average exchange rate. In addition, these banks stand to earn management fees and commissions in other areas as a result of these activities, so it is worth the risk they take. Because average-rate options have lower volatility than standard European options, they are cheaper to purchase.

Valuation of Path-Dependent Options

This section attempts to offer some insights into the valuation of two path-dependent options—the lookback and the average-rate option.

Valuing the Lookback Option. As is the case for most options, the key condition required to price the lookback option in the modern tradition is that it must be possible to hedge its risk. That is, it must be shown that the cashflow obligation(s) of the writer of a lookback call option can be exactly met by the payoff from another portfolio (a hedge portfolio). Indeed, Goldman, Sosin, and Gatto (1979) showed that such a hedge portfolio could be constructed so that the lookback option can be valued without regard to the risk premium in the underlying asset's expected return. These authors showed that when the risk-free interest rate was equal to exactly one-half the underlying asset's variance, the lookback call option is identical to the purchase of a straddle (a portfolio of puts and calls on the same assets at the same strike price) on the asset. Therefore, the writers of lookback calls can simply hedge their obligation by purchasing a straddle on the same underlying asset. Because the lookback option can be hedged, it can be valued using the risk-neutral pricing technology associated with the Black-Scholes (1973) paradigm.

Valuing an Average-Rate (Asian) Option. There are two types of average-rate or Asian options: the fixed-strike and floating-strike options. The payoff on a floating-strike Asian call option at expiration is equal to the greater of either zero or the difference between the underlying asset's terminal spot price and the average value of the asset over the life of the option—that is, $max(S - Avg_S, 0)$. It is comparable to a lookback call option for which the strike price is the average value of the underlying asset as opposed to its minimum value. Because mathematical complexities have prevented development of a closed-form analytic model (such as the Black-Scholes equation) to price such an instrument, these options must be valued with a numerical approximation technique such as Monte Carlo analysis.[3]

The value of an Asian option can never be greater than the value of a regular lookback call option, for which the strike price is the achieved minimum of the asset. Thus, the price of a regular lookback option sets an upper boundary on the average-rate option's value (because the minimum value is an extreme and the average is never equal to an extreme value unless all of the values are equal).

For fixed-strike options, the second type of average-rate option, the terminal payoff is the maximum of either zero or the difference between the average value of the underlying asset and a fixed strike price—$max(Avg_S - X, 0)$. The average can be computed using either the geometric average or the arithmetic average. Again, because of mathematical complexities no closed-form equation has been developed for pricing the average-rate option written on the arithmetic average.

Under the standard risk-neutral (Black-Scholes) pricing approach, it is assumed that the natural logarithm of stock price returns are normally distributed. In valuing an Asian option written on the geometric average of an asset's value over time, this standard assumption still holds because the product of the logarithm of stock price returns is normally distributed, and this option can be valued in closed form using the Black-Scholes approach. However, the assumption breaks down for an Asian option written on the arithmetic average because the sum of the logarithm of the stock price returns over time is no longer normally distributed. As a result, it is necessary to employ other valuation techniques for an average-rate option written on the arithmetic average of the underlying asset's price.[4]

Conclusion

It should be clear from this overview of path-dependent options that risk management is not a static field. New products and financial instruments are continuously being developed to meet new needs. While many risks can be managed with traditional hedging instruments such as standard options, futures contracts, and swaps, the rapid development of exotic options like

the path-dependent options suggests that the market for innovative risk-management products is in no way saturated. The demand for these new instruments is likely to continue growing as long as risk-management techniques using traditional hedging vehicles require close monitoring, involve fairly high commissions or management costs, and fail to reduce risks in the way desired.

The development of path-dependent options is, however, only one response to the demand for innovative risk-management instruments. In addition, because these instruments build on existing standardized derivative products, they may not serve the needs of every investor or institution. Their future development is likely to tend toward greater customizing for specific situations.

It is well known that a portfolio of existing standard products can replicate the payoffs on most of the new derivative products such as those discussed here. Indeed, this very fact allows creation of risk-free hedge portfolios for these contracts and also makes it possible to price them using the familiar risk-neutral pricing technology. However, the management and effort required for existing products to duplicate the payoff from the newer contracts tend to be too expensive an alternative for individual investors. Thus, the financial services firms that produce these new contracts add value to the market. These products have made a place for themselves because they are tailored to meet specific risk-management and investment needs.

Notes

1. The Chicago Board Options Exchange (CBOE) and the American Stock Exchange (AMEX) both trade path-dependent options known as "capped options." These options are described in detail below.

2. Such suspicions on the part of investors were not totally unwarranted. In a separate case, two Merrill Lynch vice presidents were fired for allegedly artificially driving up the price of options on a portfolio under their management on Christmas eve of 1981 in an attempt to maximize their bonus, which was tied to the portfolio's December 24, 1981, closing value (*Wall Street Journal*, January 21, 1982, 4). This example also points out the advantage of making this type of path-dependent option contingent on the average price of the underlying asset over some extended period rather than the closing price on some particular day, as the chances for artificial manipulation are greatly reduced. The growing popularity of tying conversion prices or ratios to time averages of prices in mergers seems to reflect similar concerns.

3. Monte Carlo simulating is a numerical approximation technique that can be used to compute option values by simulating the path taken by the price of the asset underlying the option over time. By simulating numerous such price paths, the technique allows one to compute the expected value or price of the option with increasing precision as the number of iterations or runs of the simulation are increased. This technique is described in detail in the forthcoming *Review* article examining the valuation of path-dependent options.

4. A more thorough discussion of the valuation or pricing of the lookback and the average-rate options, including a brief tutorial on the basic tenets of option pricing using the modern risk-neutral pricing technology pioneered by Black and Scholes (1973), will appear in the forthcoming *Review* article referred to above. The article explains how Monte Carlo analysis can be used to price Asian options written on the arithmetic average as well as how these options can be used to hedge foreign exchange risks from the viewpoint of individual investors and multinational corporations. The reader interested in the basics of option pricing is referred to Hull (1990) and Kolb (1991).

References

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (May/June 1973): 637-59.

Goldman, M. Barry, Howard Sosin, and Mary Ann Gatto. "Path-Dependent Options: Buy at the Low, Sell at the High." *Journal of Finance* 34 (December 1979): 1111-27.

Hull, John. *Options, Futures, and Other Derivative Securities*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1989.

Hunter, William C., and David W. Stowe. "Path-Dependent Options: Valuation and Applications." Federal Reserve Bank of Atlanta *Economic Review* (forthcoming 1992).

Kolb, Robert W. *Options: An Introduction*. Miami: Kolb Publishing Company, 1991.

Review Essay

Cheating the Government: The Economics of Evasion

by Frank A. Cowell.
Cambridge, Mass.: MIT Press, 1990.
267 pages. $29.95.

William C. Hunter

I am willing to bet that, as long as taxes have existed, so has tax evasion, which puts the black economy, as an activity, on a par with the oldest profession.

—Kent Matthews

*L*ike politics, religion, or sex, tax evasion is a subject that tends to evoke strong reactions—outrage, remorse, curiosity, or incredulous admiration, to name a few. And, like politics, the subject becomes a prime topic of conversation on a regular basis—around April 15 each year in the United States. Once considered unworthy of serious economic analysis because of its presumed triviality in quantitative terms or the extreme difficulty that unreliable or inaccessible data create for empirical research, the study of tax evasion has attracted increased scholarly interest in recent years.

Among economic crimes in the United States, income tax evasion is perhaps one of the most widespread, and the growth of various tax scams has undoubtedly contributed to fiscal problems not only at the federal level but also for state and local governments. The economic costs of evasion are numerous; two of the most important are lost government revenues, which must be recovered through tax programs with higher administrative costs, and the inequity between evaders and honest filers. In addition to these obvious economic costs, the mere existence of the problem has a rather disturbing implication for traditional public finance theory.

The reviewer is vice president and senior economist in charge of basic research in the Atlanta Fed's research department.

Normative and Positive
Approaches to Tax Policy

Questions related to tax policy are addressed in the fields of public economics and public finance. What types of goods should be taxed? How progressive should the income tax be? What should be the balance between taxation of commodities and the taxation of income? Such questions have occupied many of the leading economists of the last two centuries, from Adam Smith, John Stuart Mill, Francis Edgeworth, and Knut Wicksell to A.G. Pigou and Frank Ramsey. While much of the nineteenth-century literature was concerned with enunciating general principles to guide tax policy—that is, normative questions such as those listed above—modern analysis of taxation first describes the effects of taxation and then applies criteria (usually a social welfare function) to evaluate those effects. This view splits the subject into a logically prior positive side and a subsequent normative side on which value judgments are introduced. Among the former are such issues as the consequences income and wealth taxation have for risk-taking, the effects of different forms of business taxation on investment and profit distribution, the effects of tax policy on the national debt and the national savings rate and growth, and the different impacts of tax laws on various segments of the population—that is, issues related to the incidence of taxation.

Most of the modern theory of public economics rests on the notion of an all-knowing and all-powerful state (or central planner) implementing tax policy in a manner that maximizes societal welfare. This view, in turn, implies the existence of complete markets and the absence of externalities. (Externalities are costs and benefits not properly accounted for by the price system and whose existence implies that the overall welfare of society is not maximized in a market system.) However, one could argue that the existence of tax evasion is prima facie evidence of the existence of externalities. Costs associated with tax evasion, for example, are not borne exclusively by the evaders. Thus, some of the fundamental assumptions underlying much of modern public economics can be called into question. It is not surprising, then, that the subject of tax evasion and the so-called hidden or underground economy has provoked considerable academic controversy and interest.

In *Cheating the Government: The Economics of Evasion*, Frank A. Cowell provides a systematic survey of the literature addressing the economics of the hidden or underground economy and tax evasion. He also formulates and solves a series of rigorous theoretical models dealing with the various choices, decisions, and consequences of tax evasion. Cowell, a professor at the London School of Economics, is a noted authority on the economics of tax evasion and public finance and has made numerous contributions to the literature.

To the disappointment of some potential readers, *Cheating the Government* is not a handbook or users' guide to avoiding the payment of taxes. It is a scholarly and tightly argued book written primarily for the public finance specialist or professional economist. However, because it reviews empirical evidence on the magnitude of tax evasion in the United States and several other countries, it may be of interest to a more general audience as well, including those interested in tax policy from a more practical perspective.

What Is Tax Evasion and
How Significant Is It?

The income tax has made more liars out of the American people than golf has.

—Will Rogers

In my mind the simplest definition of tax evasion is that it is a deliberate attempt by an individual or company to defraud the tax authorities by giving false figures for or not declaring revenue or income, expenses, or assets for tax assessment. This description is to be distinguished from tax avoidance, which is the minimizing of tax liability by legal means. As simple as these definitions are, the boundaries separating tax evasion from tax avoidance are not as clear. This lack of clarity perhaps explains why most of the first two chapters of *Cheating the Government* is devoted to delineating the boundaries between tax evasion and tax avoidance and to defining the scope of the underground economy.

This ambiguity appears to be rooted in questions related to legality, ethics, and politics as well as economics. For example, in a purely legal sense evasion is beyond the law while avoidance is not. From a moral standpoint certain types of avoidance may be viewed as "just as bad" as evasion and deserving of the same treatment. From a political perspective evasion and avoidance may be considered two arbitrary segments of a continuum that stretches from tax planning for children's education to the fringes

of extortion. While each of these distinctions carries some force, the perspective Cowell offers is more exact. From the viewpoint of positive economics, Cowell argues that it is the certainty associated with tax avoidance that distinguishes it from tax evasion. In other words, avoidance in the strictest sense implies certainty on the part of the taxpayer at the time he makes decisions about deploying his assets and reporting to the tax authority. Evasion activities, on the other hand, involve a taxpayer's making a decision or decisions while still uncertain about his eventual tax liability.

The logic of Cowell's distinction is as follows. If the law essentially ignores a particular form of tax evasion, then, as far as the consequences to individual taxpayers are concerned, engaging in that form of evasion is no different from legitimately avoiding taxes. Conversely, if a particular avoidance scheme is actually the subject of legal doubt, or liable to substantial arbitrary penalty, then, as far as the economic consequences to the taxpayer are concerned, engaging in that scheme is equivalent to participating in manifestly illicit tax evasion.

Despite the difficulties in defining tax evasion, the consensus view appears to be that monies lost through evasion amount to as much as 2 to 10 percent of gross national product in most Western industrialized economies. For example, 5 percent is reasonable for the United Kingdom (C.V. Brown et al. 1984); the U.S. Treasury Department (1979, 1983) has suggested 6 to 8 percent for the United States; and reports of 8 to 15 percent exist for Sweden (Ingemar Hansson 1985). If these estimates contain any bias, they are likely to underestimate the phenomenon, given the difficulty of identifying evaders. In addition, the size of these estimates depends on the methodology employed.

Estimating the Extent of Tax Evasion

Measuring the extent to which an economy suffers from tax evasion is problematic because this measurement is inextricably tied to the size of the hidden or underground economy. According to Cowell, in its broadest sense the underground economy encompasses all unmeasured economic activities, that is, activities that go unreported or unmeasured by society's current techniques for monitoring economic activity. This definition includes activities excluded from the country's gross national product accounts by conven-

tion as well as those excluded because they evade the measurement process. It is important to distinguish transactions associated with tax evasion from other unmeasured but legal transactions—for example, household production and cooperative activities. Such activities are certainly part of the hidden economy but are not illegal per se, and there is no explicit attempt made to measure them or include them in official national accounts.

As noted above, the underground economy encompasses activities that, although included in the definition of measurable economic activity, escape current measurement techniques. Unreported income and certain fringe benefits as well as the production of illegal goods or services fall into this category. Clearly, using any given definition of tax evasion, problems arise because a certain number of economic transactions will not fit neatly into any category and because, under the existing legal or social structure, activities taking place in one sector of the economy (for example, the underground economy) will produce relevant effects on the other (the measured economy). A simple example given in *Cheating the Government* illustrates how nebulous these boundaries can be. Suppose that a homeowner decides to repaint certain rooms in her house. She can employ a painter and pay cash on which the taxes are evaded, or she may undertake the painting herself. The former transaction falls in the realm of the illegal side of the underground economy, while the latter is clearly part of the legal underground economy. The homeowner's choice between the two transactions can be modeled as the outcome of a standard optimization problem in which the homeowner chooses the transaction that yields the maximum utility subject to the constraints imposed by household budget and existing social and legal institutions.

In the second chapter of *Cheating the Government* Cowell reviews the methods, both direct and indirect, that economists employ in estimating the size of the underground economy. Direct methods include projections based on the results of intensive audits of individual taxpayers (or samples of taxpayers) and projections based on the results of surveys of individuals' economic activities and attitudes. These methods concentrate on the tax evasion aspect of the underground economy and can provide reliable though conservative estimates of this sector's size.[1]

In the United States the most notable attempts to measure the extent of unreported income by extrapolating from the results of intensive taxpayer audits are conducted by the Internal Revenue Service as part of

its Taxpayer Compliance Measurement Program. This program has been in effect since 1963 and supplements the quick checks made on all tax returns and the more detailed audits conducted on suspicious returns. The Taxpayer Compliance Measurement Program involves extensive auditing of approximately 50,000 randomly chosen taxpayers whose tax returns are matched against information on interest payments and dividends as reported by businesses and other organizations as well as individuals who report substantial income related to services rendered. An obvious problem with such tax compliance programs is that they do not include individuals who simply do not file tax returns.[2] A conservative estimate for 1981 derived using these methods indicates that unreported income amounted to 8 percent of GNP.[3]

Indirect methods of estimating the size of the underground economy may use as a gauge certain discrepancies between various economic quantities or indicators or may be based on the assumption that there are stable relationships among various aggregate monetary measures. The expenditure-income discrepancy method relies on the assumption that, while some income earned by individuals will go unreported or underreported, much of it will eventually show up in the form of expenditures. Evaluating the difference between the amount of income in the economy's national income accounts and income estimates based on adjusted tax returns yields an estimate of the size of the underground economy and tax evasion. The U.S. Commerce Department's Bureau of Economic Analysis estimated the expenditure-income discrepancy to be about 5.5 percent for the United States in 1968. Estimates for other countries based on this method are 4.7 percent for Sweden in 1978, 6 percent for Denmark in 1977, 9 percent for West Germany in 1968, about 20 percent for Belgium in 1970, 23 percent for France in 1965, and 10 percent for employees and 23 percent for self-employed individuals in Italy in 1980.

Some indirect methods for estimating the extent of tax evasion and the size of the underground economy are derived from the analysis of the economy's monetary aggregates. The currency denomination approach is based on the notion that there is a positive and direct relationship between the size of the underground economy and the number of large-denomination bills in circulation—in other words, as the size of the underground economy increases, so does the number of large bills required to facilitate unreported and untaxed payments. For example, in the United States from 1960 to 1970, while the supply of small-denomination

bills ($1 to $10) increased by 37 percent, the stock of large-denomination bills rose by more than 75 percent.[4] From 1966 to mid-1978 the stock of $100 bills increased by more that 250 percent. Similar results are reported for other developed countries. Interesting as this type of estimate is, this approach is not taken very seriously because it has many shortcomings. For example, inflation could account for the growth in large-denomination bills.

A somewhat more reliable indirect method is the currency-to-demand-deposit-ratio approach. This approach assumes that the size of the underground economy can be garnered from the ratio of outstanding currency to demand deposits held in the commercial banking system. It relies on there being a stable and predictable relationship between the cash held by the private sector and the total monetary base. By fixing a date at which the underground economy is assumed to be virtually nonexistent, the size of the underground economy can be estimated by observing the growth of currency in excess of the monetary base. This approach was first used by Phillip Cagan (1958) and further developed by Peter M. Gutmann (1977) and Edgar Feige (1989).

Two other methods of estimating the size of the underground economy are the labor-market and the causal-modeling approaches. The labor-market approach holds that a low labor force participation rate can be used to explain the size of the underground economy if the rate is out of line with some base period or with participation rates in similarly situated countries. The causal- (or soft-) modeling approach essentially identifies those variables that should be correlated with the size of the underground economy and feeds various values of these variables into larger structural models of the economy in an attempt to provide reasonable estimates.

All of these methods suffer from the fact that they rely on restrictive and, in many cases, untestable assumptions, such as the existence of a unique base period during which the size of the underground economy was negligible or the notion that the velocity of circulation of cash in the underground economy is the same as the velocity of circulation of money in the regular economy. A more glaring policy-related weakness associated with these methods is that, although they use evidence from the demand side of the economy, they are often used to draw conclusions concerning the supply side of the underground economy—to estimate, for example, the extent to which unemployed individuals are active in the underground economy or the amount of tax evasion that occurs.

Why Do People Try to Evade Taxes?

The avoidance of taxes is the only pursuit that still carries any reward.

—John Maynard Keynes

In Chapters 3 through 6 of *Cheating the Government* Cowell provides answers to the above question by examining formal models of the rational taxpayer, the taxing authority, and the production side of the underground economy. Examining the taxpayer, the author uses an approach similar to that in the public economics literature, which assumes that the rational taxpayer acts to maximize expected utility based on his or her perceptions of the probabilities that evasion will be detected and penalized, the nature of the penalty, and the taxpayer's income and degree of risk aversion. The potential gain from evasion is the amount of tax liability not paid; the potential costs are the penalties (which might be monetary, criminal, psychological, or some combination of these) that the taxpayer would suffer if evasion is detected and successfully prosecuted.

Cowell's answers to the question of why taxpayers cheat are numerous but fairly intuitive and have to do with the trade-off the taxpayer makes between the utility received from cheating and not being caught and the disutility associated with cheating and being caught. For example, Cowell's model predicts that more risk-averse individuals tend to evade less and those with higher incomes tend to evade more and that the larger the general economy is relative to the individual taxpayer, the more likely the individual is to attempt evasion. As would be expected, Cowell's analysis also indicates that increasing the probability of detection reduces the incidence of evasion and that, under certain conditions, tax cuts without any changes in the penalties for evasion will tend to increase the amount of evasion while under other conditions increases in the progressiveness of taxes can reduce the amount of tax evasion.

The results obtained for taxpayers in the aggregate are more complicated and depend on the relationships among the parameters in the model. However, at the aggregate level Cowell's model predicts that evasion will vary considerably with the type of income earned and individual characteristics (for instance, individuals have a relatively low propensity for underreporting income when compared with, say, farmers and others reporting business income; married people evade more than single persons; and younger people

evade more than older persons). In addition, the model predicts that higher disposable income combined with lax enforcement efforts will result in more evasion and that evasion generally increases with increases in the tax rate—that is, there is less compliance among taxpayers in brackets with higher marginal tax rates.[5] After reading Cowell's discussion of the reasons people evade taxes, one is struck with the idea that if evasion rises with increases in the tax rate, then lowering tax rates may actually increase the total tax revenues collected by the government if evasion is reduced significantly. This type of thinking, combined with the supply-side notion that lower taxes also significantly stimulate business activity, undoubtedly informed the tax cuts of the early 1980s.

The examples discussed here are only a few of many interesting insights provided by Cowell's analyses of the rational taxpayer, the taxing authority, and the interactions between regular and underground economies. Although Cowell's theoretical results are occasionally ambiguous and require restrictive assumptions to produce predictions that conform to observed empirical evidence, he makes it clear that tax policy can have wide-ranging and long-lasting effects on the entire gamut of economic activity structure, from businesses' choices of capital investment to individual citizens' occupational choices. Tax enforcement policy can have similar effects.

Stamping Out Tax Evasion

If you can't do the time, don't do the crime.

—Robert Blake as *Beretta*

If you can't pay the fine, don't do the crime.

—Anonymous

Chapters 7 and 8 of *Cheating the Government* address the public policy issues related to tax evasion and the underground economy. Within the context of the basic model developed in chapter 4, Cowell shows that tax evasion can be eliminated through the proper use of legal penalties and fines. However, to his credit, Cowell is quick to point out that the "model" solution is not very appealing or practical. In particular, given such considerations as the real resource costs associated with enforcing compliance at the levels necessary to eliminate evasion and the unjust nature of the fines that would be required, the solution of increased fines and stronger legal penalties is simplistic. Furthermore,

reducing the level of evasion to zero too quickly could have adverse effects on national income and the standard of living. These considerations point out the dangers involved in using such a simplified model to make public policy prescriptions.

Cowell next explores ways in which tax administration (enforcement and collection) can be improved. He develops several simple rules for efficient tax administration using the analytical framework developed in earlier chapters. As is the case with most resource allocation problems in economics, the simplest rule for efficient tax administration balances the costs and benefits associated with carrying out enforcement and collection activities. For example, antievasion instruments should be set in such a way that the expected revenue raised by a marginal change in each instrument should exactly equal the marginal cost of changing that instrument. Using this rule, if the marginal benefit of a particular instrument happens to exceed its marginal cost, then using that instrument more intensively should increase benefits. However, this logic presupposes that the "benefit" to be achieved by a tighter control of tax evasion is to be understood purely in terms of increasing the amount of revenue collected, that is, tax farming. It also presupposes consensus in identifying the costs associated with eliminating evasion.

Again, Cowell does not advocate that the prescription derived from this analytical model be followed blindly. He stresses that such a simple rule neglects some rather important issues of economic policy. In considering taxation policies the government as a whole is presumably concerned with some notion of social welfare that is broader than just the amount of revenue it takes in. It should also take into account what the revenue is to be used for, who benefits from an enforcement policy, and resource costs other than those required to police the tax system. Issues such as social justice and efficiency of the production sector should be included in the policy agenda; otherwise, narrow-minded pursuit of a single objective is likely to lead to absurdities. Accordingly, Cowell argues, the rules for the design of enforcement and collection policy must be put on firm footing.

The remainder of chapter 8 is devoted to establishing this foundation for enforcement and collection policy and then making recommendations for improving current tax administration policies. Cowell considers governmental budget, information, and political constraints and ponders the appropriate objective function for the government along with the system responses it is likely to engender. Through careful anal-

ysis of these factors the author is able to offer several practical suggestions for improving the enforcement and collection aspects of tax administration. Switching from voluntary reporting systems to withholding systems, shifting from direct to indirect taxation, allowing penalties for evasion to vary more directly with the amount of tax evaded, and moving toward nonrandom audit policies are just a few of Cowell's prescriptions.

In the book's final chapter Cowell reiterates that tax evasion is more than just a curiosity. Given the empirical evidence in the public finance and public economics literature and the increasing percentage of income claimed as taxes in many countries, there is little doubt that the subject is worthy of serious economic analysis. The questions and issues surrounding tax evasion lie at the heart of public finance and public economics, casting doubt on some conventional wisdom on the one hand and raising new and interesting paradigms on the other.

Some Closing Thoughts

In *Cheating the Government*, Frank Cowell provides a comprehensive survey of the tax evasion literature and several insightful extensions of this literature. While the book will certainly become a basic reference work on the subject for some time, it is lacking in one area. Cowell does not examine in any detail the subject of revenue enhancement, that is, innovative ways for generating revenue to replace that lost to tax evasion.

Beleaguered by declining tax revenues and mounting expenditures, many state and local governments have sought alternative and novel revenue sources. State lotteries and legalized gambling are two responses to this need to generate additional revenue. One approach that has been used by about twenty-eight states since 1981 is a tax amnesty. Like amnesties for overdue library books or unpaid parking tickets, tax amnesties give individuals an opportunity to pay previously unpaid taxes without being subject to normal penalties or prosecution. While many states have generated large amounts of revenues with these programs, other states have been less successful. For example, New York collected $401 million and California, Illinois, and Michigan each garnered more than $100 million. In contrast, amnesties in North Dakota, Idaho, Texas, Kansas, and Missouri yielded less than $1 million each.

Besides the fact that their potential to generate revenues is uncertain, tax amnesties are also controversial revenue tools because of concern about their effect on voluntary compliance. Advocates of amnesties argue that a one-time amnesty may increase future compliance if it is accompanied by greater expenditures for future enforcement and stronger penalties for evaders. Critics, on the other hand, contend that the long-run consequences could be less voluntary compliance. Given the growing interest in amnesties in several states, thorough analysis of the long-term benefits and costs of tax amnesties is clearly called for.

Notes

1. Typically, in making such projections, the tax authority can reliably report only on those violations of tax laws over which it has direct jurisdiction. Surveys, on the other hand, are subject to the problems of nonresponse, evasiveness, and misrepresentation because the respondents feel compelled to supply answers that are consistent with filed tax forms. Thus, these projections tend to be biased downwards.
2. The IRS currently combines the information obtained from its Taxpayer Compliance Measurement Program with estimates of unreported income made by the United States General Accounting Office to arrive at more accurate estimates.

In 1979 the GAO estimated that about 5 million individuals failed to file tax returns.
3. These methods have been applied extensively in other countries. See, for example, Brown et al. (1984) and Isachsen and Strøm (1985).
4. See "The Growing Appetite for Cash," Federal Reserve Bank of Chicago *Business Conditions* (April 1971): 16.
5. See, for example, Feinstein (1991) for a recent detailed econometric analysis of income tax evasion and its detection based on the data contained in the Internal Revenue Service Taxpayer Compliance Measurement Program.

References

Brown, C.V., E.G. Levin, P.J. Rosa, and D.T. Ulph. "Tax Evasion and Avoidance on Earned Income: Some Survey Evidence." *Fiscal Studies* 5 (1984): 1-22.

Cagan, Phillip. "The Demand for Currency Relative to Total Money Supply." *Journal of Political Economy* 66 (1958): 303-29.

Feige, Edgar. *The Underground Economies: Tax Evasion and Information Distortion*. Cambridge, Eng.: Cambridge University Press, 1989.

Feinstein, Jonathan S. "An Econometric Analysis of Income Tax Evasion and Its Detection." *Rand Journal of Economics* 22 (Spring 1991): 14-25.

Guttman, Peter M. "The Subterranean Economy." *Financial Analysts Journal* 33 (January/February 1977): 24-27.

Hansson, Ingemar. "Tax Evasion and Government Policy." In *The Economics of the Shadow Economy*, edited by A. Wenig and W. Gaertner. Berlin: Springer-Verlag, 1985.

Isachsen, Arne J., and S. Strøm. "The Size and Growth of the Hidden Economy in Norway." *Review of Income and Wealth* 31 (1985): 21-38.

U.S. Department of the Treasury. Internal Revenue Service. *Estimates of Income Unreported on Individual Tax Returns*. Washington, D.C.: 1979, 1983.

Federal
Reserve
Bank of
Atlanta

Public Affairs Department
104 Marietta Street, N.W.
Atlanta, Georgia 30303-2713
(4 0 4) 5 2 1 – 8 0 2 0

R0000000002951-0
MR. HERB TAYLOR
FED RESERVE BANK OF PHILADELPHIA
10 INDEPENDENCE
PHILADELPHIA, PA 19106-1574