ARTICLES

Global Social Security: How the Rules are Changing in Certain Countries
Robert J. Myers

On Becoming a Cost Effective Company
Robert D. Shapiro and Barton H. Clennon

The Process of Pension Forecasting
Michael Sze

Cost-Based Pricing of Individual Automobile Risk Transfer: Car-Mile Exposure Unit Analysis
Patrick Butler

Discussion
Ruy A. Cardoso
Richard G. Woll
Patrick Butler's Reply

A New Approach to Modeling Excess Mortality
Peter D. England and Steven Haberman

The Small Plan Audit Program: The Opinions of the Court
Arnold F. Shapiro

BOOK REVIEWS


The editor communicates the recommendation to the author(s) along with copies of the referee reports. The entire process is expected to take three to four months.

The Journal welcomes articles providing new (or improved) strategies or techniques that can be used by practicing actuaries. In this regard, one of the goals of the Journal is to improve communication between the practicing and the academic actuarial communities.

The Journal publishes articles in a wide variety of formats, including technical papers, commentaries/opinions, discussions, essays, book reviews, and letters. The technical papers published in the Journal are neither abstract nor esoteric; they are practical and readable. Topics suitable for this journal include the following areas:

| AIDS             | annuity products | asset-liability matching | cash-flow testing | casualty ratemaking | credibility theory | credit insurance | disability insurance | expense analysis | experience studies | FASB issues | financial reporting | group insurance | health insurance | individual risk taking | insurance regulations | international issues | investments | liability insurance | loss reserves | marketing | pensions | pricing issues | product development | reinsuranc | reserving issues | risk-based capital | risk theory | social insurance | solvency issues | taxation | valuation issues | workers’ compensation |
|-------------------|------------------|--------------------------|-------------------|--------------------|--------------------|------------------|-------------------|----------------------|-------------------|-------------------|-------------|-------------------|-------------------|-------------------|----------------------|----------------------|----------------------|-----------------|-----------------|-------------|-----------------|-------------------|-----------------|------------------|-------------------|-----------------|-----------------|-----------------|-----------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|-----------------|-----------------|-------------------|-------------------|

A paper submitted first is screened by the editor for suitability. If it is deemed suitable, copies are sent to at least three independent referees. The name of the author(s) of the paper under consideration is anonymous to the referees, and the identities of referees are not revealed to the author(s).

The referee reports are reviewed for content, originality, and clarity of exposition. On the basis of the referee reports, the editor makes one of the following recommendations: (1) accept unconditionally, (2) accept subject to minor revisions, (3) accept contingent on substantive revisions, (4) resubmit, return to the author(s) for major revisions and subsequent resubmission, or (5) reject.

The editor communicates the recommendation to the author(s) along with copies of the referee reports. The entire process is expected to take three to four months.

See back cover for instructions to authors.
Global Social Security:
How the Rules are Changing in Certain Countries

Robert J. Myers*

Abstract

Social security programs (i.e., national pension systems) differ widely between countries. This is only natural, and desirable, because of varying social and economic conditions and philosophies.

This paper discusses some of the general worldwide trends, such as the equal treatment of men and women, increases in the normal retirement age, projection of future costs, and the different philosophies of social security. Some of the interesting and unique changes recently made in selected countries—Canada, Chile, Eastern European countries, Germany, Japan, People's Republic of China, Saudi Arabia, Union of Soviet Socialist Republics, United Kingdom, and United States—are described.

Key words and phrases: global social security, social security, national pension systems, philosophies of social security

1 Introduction

There have been several interesting and significant developments that recently have occurred in social security around the world. Some of these suggest worldwide trends, whereas others are unique to particular nations. This paper describes such developments in several selected countries.

The term social security as used here means only the limited concept of a national pension system. It does not include programs such as unemployment insurance, family allowances, workers' compensation, and health care that some persons consider to be branches of social security.

2 Worldwide Major Developments

In recent years two major developments in social security have occurred in most countries. The first one is equal treatment for men...
and women. This is an easy concept to understand. If a higher retirement age applies for men than for women (often a difference of five years), for example, this is not equal treatment. Also, many differences occur where women have been treated less favorably than men: survivor benefits, for example, have been available in some countries for male workers, but not for female workers. Many technical difficulties arise, however, in implementing equal treatment.

The second major development in many countries is the likely future financing problem as the population ages; in other words, as there are relatively more persons at retirement age compared to persons of working age. This growing proportion of older persons creates (or can create) financial problems, especially if a country does not recognize that financing problems are likely to occur in the future.

Some years ago, few countries did any serious forecasting of what their social security programs would cost 20, 30, or even 75 years hence. Many countries merely looked ahead a year or two. This worked well for some years, but as the aging of the population has continued, the financial burden has become heavier. In contrast, some countries (such as the United States) have made long-range projections for many years.

Projecting 50 to 75 years into the future cannot be done with great precision, any more than one can predict with precision what the weather will be in several weeks. But just as with the weather, one does know that, if it is summer and winter is coming in about six months, it will be colder then. One may not know exactly how much colder, though. Long-range projections have given many countries some indication of the problems that will be coming. In recent years more countries have become concerned about the long-range future costs of their social security programs.

3 Role of Visiting Experts

Next I will discuss some of the most interesting specific developments in several countries with which I am familiar. Obviously, one article cannot describe what is happening in all countries!

When a person travels to another country as a consultant in the field of social security (or, for that matter, any other field), he or she should not use what is done in his or her own country as an absolute guide for what other countries ought to do. Instead, technical experts should consider what the particular situation is in a country as compared not only with their native land, but also with countries throughout the world. What operates well and is desirable in one country frequently may have just the opposite outcome in another
country. There is no one perfect way of doing things; there are many different alternatives. The choice of which alternative to take is often not merely financial, actuarial, or economic, but is also dependent upon the political or psychological characteristics of the country.

4 United Kingdom

Now let us take a tour around the world, going eastward from the Americas. The first country I will discuss is the United Kingdom, which faces a serious problem concerning equal treatment by sex. This problem has been especially acute since the United Kingdom joined the European Economic Community (E.E.C.). The U.K. has had a five year differential in the minimum retirement ages for men and women in its social security program (60 for women and 65 for men). The benefits for women with similar earnings records often are higher in order to make up for the fact that their contribution period or service period is shorter. The E.E.C., however, believes that there should be equal treatment of men and women in all respects—social security, pensions, and so forth.

The U.K. has a dilemma because private pension plans must have equal treatment; if they do not, legal suits can be brought in E.E.C. courts. On the other hand, the E.E.C. doctrine on equal treatment does not control social security systems completely. At this time an employer in the U.K. with a private pension plan must provide equal treatment; if women can retire at age 60 with a certain amount of pension, men must have the same amount at that age. At the same time, however, the social security system does not pay the same benefit to men as to women, especially at ages 60 to 64. The employer cannot bridge the gap by providing a temporary benefit to men to equalize treatment in the aggregate between men and women because it would be unfair discrimination against women; men would have larger benefits from the private pension plan!

The real solution to this problem is to have the social security system also provide equal treatment. The U.K. government is struggling with this matter. One difficulty with equalizing retirement ages between men and women by lowering the age for men is the greatly increased cost of the program. But if the age for women is raised, many female workers will be extremely dissatisfied, especially those now near the current retirement age. In the end, however, this equalization must be done. Probably the best way to equalize treatment of men and women is not suddenly, but with a
gradual transition. In any event, the U.K. has a difficult problem in this area.

Another interesting development in the U.K. is connected to individuals opting out of the social security system. The U.K. system is composed of two parts: a flat benefit and an earnings-related benefit. For many years, employers have been able to opt out of the earnings-related benefit portion if they provide at least equivalent benefits. Although this makes the system complicated, it has been working reasonably well. Over the past two or three years, however, changes in the system have allowed persons in a plan that had opted out of the earnings-related benefit portion to opt out of the plan individually if they provide their own retirement protection. Also, persons whose employers do not opt out can opt out individually. Individual opting-out is undesirable, because it will be difficult to prevent adverse selection and the resulting increased costs. At the same time, the principle of social solidarity is violated.

5 Former Soviet Union

The next country is the former Soviet Union, which has the same problem as the United Kingdom: unequal retirement ages for men and women (namely, 55 for women and 60 for men). A Russian colleague of mine informed me that this is a great concern. Their experts know that they should have equal ages by sex, but this is difficult to achieve from a political standpoint. With all of the other problems facing the former Soviet Union, however, this one is undoubtedly not high on the list of priorities.

Another problem in the former Soviet Union is that pensions are low. For many years I attended international conferences on social security where Soviet delegates would proclaim that they had the most wonderful social security system. They asserted that it took care of all the needs of all their people and that it was paid for entirely by employing entities and by government and not at all by workers.

The level of benefits of the social security program relative to earnings in the Soviet Union a few years ago was close to that in the United States. As in the United States, the benefits are graded, i.e., relatively higher for low income persons and relatively lower for high income persons. For example, for a worker in the United States with average earnings over the working lifetime, the benefit is about 42 percent of final wages. For the low paid worker, the benefit is 55 percent to 60 percent of final wages. For the highest paid worker (up to the maximum earnings considered for benefit purposes), the benefit is 25 percent to 27 percent of final wages.
In the former Soviet Union, the minimum retirement ages are lower than in the U.S. Their level of total retirement income is relatively low, however, because the social security benefits are the only source of retirement income in most cases. In contrast, many persons in the U.S. have private pensions in addition to Social Security benefits, as well as more private savings, home ownership, and so forth. The total retirement-benefit level in the former Soviet Union is low, and its policy makers are concerned about the situation. With the recent horrendous inflation, the purchasing value of benefits has dropped sharply despite month by month ad hoc adjustments (which essentially merely raise the minimum pension so that virtually all beneficiaries receive the same amount).

A surprising development occurred in the social security field in the Soviet Union in 1988. The former Soviet Union then had only one insurance company, Gosstrakh, which was owned by the government (although some individually owned companies now are being established). Gosstrakh sells insurance policies of the standard forms that life insurance companies in any country sell, although it tends to specialize in short-term endowment policies of five to ten years. These policies are sold by agents, as in other countries. The premium rates are determined actuarially, so that all policyholders are paid an equitable amount, and the system costs the government nothing. The government probably even makes a profit on it.

In 1988, Gosstrakh began writing individual deferred-annuity policies, under which individuals could buy a certain unit of monthly pension (such as ten rubles), beginning at age 60 for men and age 55 for women. Although these policies were sold by agents, the premiums were collected through payroll deduction. This was unlike their life insurance policies, under which agents usually came to the home to collect premiums.

The basic reason for this new plan, as stated in the decree that established it, was that social security benefits were too low, particularly for workers at average and higher earnings levels. In this way, those in this economic category could provide more adequate retirement incomes for themselves on a voluntary basis.

The premium rates, unlike those for the life insurance policies, were not established on an actuarial basis. Rather, the premium rates reflected a considerable government subsidy. Thus, this plan involved a government policy to increase individual retirement income, but to have individuals partially pay for it directly.

To an actuary, it seems strange that the same premium rate was charged for men and women for a given amount of pension that was deferred for a prescribed number of years in spite of the fact that
women live longer. Further, the premium rate for a woman age $x$ who received the benefit 30 years later (at age $x + 30$) was the same as the premium rate for a man age $x + 5$ who did not receive the benefit until age $x + 35$. A double action was present, which resulted in bargain rates to women because of their favorable mortality and because of the earlier age at which they received the pension.

The rates were graded actuarially by age at issue, however. If one bought a benefit of ten rubles a month at retirement age, the premium was much higher if the policy were bought a short time before retirement age was reached than if a longer period of deferment was involved.

Considerable interest in the new voluntary-annuity program was expressed when it began operating in 1988. A reported 400,000 policies were sold in the first year. By 1989, when extensive liberalizations in the social security benefits were proposed by the government, however, interest in the voluntary annuities plummeted. Thus, most of the policies were allowed to lapse, and few new policies were written. An interesting (and amazing) development apparently came to an end and is unlikely to be resurrected, considering the political and economic upheaval in the Soviet Union in 1990 to 1991. (This upheaval also made existing policies virtually worthless as a result of inflation.)

The dissolution of the Soviet Union into separate independent nations has created many problems in the social security field. Whether each of the nations will establish new systems, how such systems will be funded, how the new nations will divide the old system and its assets, and how they will deal with persons who worked in different former republics are unresolved questions facing the new countries of the former Soviet Union.

6 Germany

Germany is experiencing just the opposite situation. But the reunification of Germany, essentially a merger of East Germany into West Germany, presents many of the same problems in the social security area. West Germany essentially has absorbed the East Germans into their social security system and will pay the extra costs involved. The system for the reunified Germany will be much like (if not entirely the same as) the previous system for West Germany. Nonetheless, some transitional problems will be present, particularly in areas where the East German program provisions were more liberal.
7 Eastern European Countries

The Eastern European countries have problems with their social security programs that are similar, in some ways, to those in the former Soviet Union, as well as some uniquely different problems. Their retirement ages vary by sex and are also very low, which results in high contribution rates. Unlike the former Soviet Union, their benefit levels are high, further resulting in high contribution rates. Their disability experience is high, in part due to loose administration. Coverage compliance has deteriorated as the societies in Eastern European countries have become freer.

Some economic planners within the Eastern European countries—as well as visiting experts from other nations—seek to privatize, in whole or in part, their social security programs along the Chilean line (as discussed later). At the same time, they would like to turn over to the new system some of the assets of former nationalized industries and companies. From another point of view, however, experienced administrators of the social security programs seem to believe that solutions to their problems can be found within the traditional framework of social insurance.

In any event, it seems likely that the level of benefits under some of the Eastern European systems will be lowered somewhat. At the same time, private pension plans (along traditional lines, including private sector investments) are expected to develop.

8 Saudi Arabia

Let us next go south and east to Saudi Arabia. This country has a traditional social insurance system, with contribution rates of 8 percent from the employer and 5 percent from the employee. The pension is related to the individual’s most recent salary. Initially there was a very liberal vesting provision, so that persons who worked just a few years and then left the country were eligible for a partial pension payable when retirement age was reached.

Many foreign workers come to Saudi Arabia for short periods. These workers are not only from the United States and Europe, but from many other countries throughout the world, such as Korea and the Philippines. In many ways, this liberal treatment for foreign workers said, in essence, “You’ll get a partial retirement pension when you reach retirement age, which will be sent to you in your home country, even though you have been out of Saudi Arabia for some years.”
Several years ago, the law was changed. Persons who are not living in Saudi Arabia at the time when they reach retirement age no longer can obtain these vested pensions. Instead, they receive only a refund of employee contributions without interest. This has helped the financing of the Saudi system greatly, because all employer contributions and investment earnings on employee contributions remain within the system. One difficulty in the Saudi system is tracking the location of foreign workers over time. As with many social security systems, when individuals seek benefits, they must go to the system and ask for them. There are many persons who have worked in Saudi Arabia over the last 20 or 30 years who may forget that they have vested pensions coming from the Saudi system when they reach retirement age. It is unlikely that they have heard that the only thing that they can receive is the refund of their contributions.

9 People's Republic of China

Our next stop is China. For the 90 percent of its huge population in rural areas, no national pension system or social security program exists. For workers in industry, commerce, and government, however, legislation has required each establishment to set up a pension plan of a more or less standard type for some years. For example, a steel mill must have a pension plan for its employees. These pension plans usually have a retirement age of 60 for men and 55 for women—again, the problem of unequal treatment by sex—and they pay benefits of about 70 percent of final wages for a lifetime of employment. The plans are financed entirely by the employing establishment, completely on a pay-as-you-go (or current-cost) basis. In other words, there has been no funding (or even establishment of reserves) for persons who currently are retired. Another problem is that individuals are required to be in service when they reach retirement age. Thus, if they move from job to job, almost all pension rights are lost.

In the past five years, the Chinese government has been more concerned about matters relating to economic development. The government has decided that the previous employment system (under which once a person was hired for a job, it was a lifetime one) is not desirable. It now believes that there could be more productivity if there were freedom of movement from one type of employment to another. But the difficulty with this change is that pensions often would not be available because of the lack of vesting.

Another economic development problem is that companies or establishments that have been operating for many years have a relatively high pension cost because current pensions are paid with cur-
rent income. A similar establishment that has just begun operations has no current pension costs and, therefore, can produce at a much lower cost. Thus, the older establishments are at an economic disadvantage.

The Chinese government is concerned about how pay-as-you-go financing of private pension plans affects their economic development. As a result, government officials have been thinking about having a national system to equalize the cost between new and old establishments. Naturally, the new establishments (and the provinces where the establishments are mostly new ones) prefer the status quo because it results in lower costs for them. They do not want to share the higher pension costs of Shanghai or Beijing. This is currently a difficult political, as well as technical, problem in China.

10 Japan

Next let us turn to Japan, which currently has the lowest mortality in the world (in other words, the greatest longevity). This, in turn, means high social security and pension costs. The Japanese government has recognized for some years this coming trend and gradually has increased the minimum retirement ages.

Japan has two national pension systems. One provides flat benefits for the entire general population: not only employees, but also self-employed persons (farmers, operators of small businesses, and so forth). The other is an earnings-related program that applies in manufacturing and commercial industries. In the flat-benefit plan, the minimum retirement age has been increased to 65 for the normal pension for both men and women, but individuals may retire as early as age 60 and receive a reduced pension. On the other hand, persons can retire later, up to age 70, and receive an increased pension.

In the earnings-related plan, the retirement ages at one time were 60 for men and 55 for women, but they are being increased by five years for women on a gradual transitional basis (reaching age 60 for those born after April 1, 1941), which eventually will solve the problem of unequal treatment by sex. The Japanese are concerned about the relatively low retirement ages; some persons in the government want to increase the age for both men and women to 65 in order to solve the problem of high cost that will occur as the population ages. Although the government wants to make this change, the situation is difficult politically. When this change is made, it will be phased in gradually; at the moment, however, it has been put aside until some more propitious time when the government hopes
there will be fewer complaints from both men and women about raising the retirement age.

Another interesting feature in the Japanese system—one that is surprising and one that the authorities now have become aware of—is the factors that are used to adjust benefit amounts for early and late retirement in the flat-benefit plan. Decreases are made because of early retirement and increases because of late retirement.

Despite the technical and actuarial expertise available in Japan, somebody erred when the adjustment factors were established. For a person retiring at age 60, the reduction for not waiting until age 65 to receive benefits should be generally about 30 percent. In the Japanese plan, however, the reduction is 42 percent, a very bad deal from an actuarial standpoint. Thus, if persons can avoid filing for benefits at age 60 and wait until age 65, they are in a much better financial position. Also, rather surprisingly, there is no graduation in the provision. The factor depends on integral years of age; in other words, there is only one reduction factor applicable if retirement occurs between ages 60 and 61, but another smaller factor—35 percent—applies for retirement at age 61. It would seem more reasonable if the reduction factors moved smoothly from age 60 to age 65, with monthly changes.

The beneficiaries involved are aware of this situation; almost everybody takes the benefit at an exact age. The surprising thing is that so many persons take benefits at age 60. Some undoubtedly have to because they do not have other resources, but there are many others who would not have to take benefits. Many persons are disadvantaged by not realizing that a bad deal exists!

At the other end of the retirement band, if instead of taking the benefit at age 65, persons wait until age 70, the actuarial increase should be 40 percent to 50 percent. In this system, however, this differential is 80 percent. Anybody in good health who had the advice of an actuary would not take the pension until age 70! In actual experience, very few persons do.

11 Trust Territory of the Pacific Islands

Next let us go to the Trust Territory of the Pacific Islands, otherwise known as Micronesia which has a population of about 200,000. This is a group of islands in the Central Pacific that the United States received as a trust from the United Nations after World War II. In 1986, the Trust Territory was divided into four parts, three of which are now independent nations—the Republic of Palau, the Federated States of Micronesia, and the Republic of the Marshall
Islands. The fourth part, the Northern Mariana Islands, voted to become part of the United States (just as is Guam, which in essence is the Southern Mariana Islands).

The Trust Territory established a social security system in 1967 at the request of the United Nations, which held that a good trustee should develop a social security system for such a territory. (The author worked on this project, and the system was established and operated successfully thereafter.)

When the Trust Territory was divided, a unique problem arose: how to divide a social security system equitably among different geographical regions. A subdivision was made, and each entity received an equitable share of the assets (and of the future liabilities, too). The Northern Mariana Islands system merely joined the U.S. system, and credit was given for all prior service as though it had been performed in the continental United States. The three new independent countries started with the existing system, but undoubtedly they will modify it in the future. Many persons there think that a retirement age of 60 is too high and they want to lower it. These individuals may not realize that the long-run high costs of such a move will be difficult to bear.

12 Canada

Next we come to Canada. One small change made in 1991 greatly affected the underlying philosophy of its social security system. Canada, like the United Kingdom, has two plans. One is called *Old Age Security*, under which a flat amount is payable to every person in the country age 65 or older who meets certain residence and citizenship requirements. The other is an earnings-related system, called the *Canada/Quebec Pension Plan*. The combination of these two plans produces a weighted-benefits structure, just as prevails in the U.S. system. With the flat benefit and an earnings-related benefit, lower paid persons receive relatively higher benefits than do higher paid ones. The combined level of benefits in Canada is about the same as that in the U.S.

The small change in Canadian policy was made considering only budget effects and not the long-range social effects. The government introduced what some persons refer to as the *Claw-Back*. This is analogous to a lobster clawing money back!

Under this provision, individuals with moderately high income, roughly C$50,000 a year or more, must return part of the flat pension. This provision is to be phased in over several years. After some years, the benefit under the Old Age Security system will not be
available to the highest income Canadians. The income limit at which this applies is only partially indexed; as time goes by, more persons will be affected by the provision. The system will become more a public assistance (or social assistance) system instead of a social insurance or demogrant program. This has been a significant change in the philosophy of the Canadian system. It is not clear whether the change was intentional.

13 Chile

The new Chilean system, which was established in 1981, involves privatization and individual defined-contribution accounts that are determined in real terms (i.e., indexed for inflation). It represents one of the most interesting and important developments in social security in the last decade or so. Many countries around the world—not just North, Central, and South American countries, but also some European countries—are interested in this emerging pension system.

Many observers do not realize that the system involves more than privatized individual accounts. The government also must make mammoth transfer payments from general revenues to meet the cost of prior service credits and large minimum-benefit guarantees. Further, much of the investments of the private funds are in government bonds, which probably were issued to meet the foregoing costs—a circular effect! Although this large cost to the government might be acceptable in Chile (which would have had equally high, or even larger, costs under its previous system), this might not be acceptable in other countries.

The Chilean system is now 12 years old, and it seems to be operating well. A cautious actuary must say, “Twelve years is a short time in the life of any sort of pension plan.” Not that any catastrophe is likely to occur, but its experience may not be as favorable as its supporters anticipate. In particular, the real interest rate earned by the various privatized funds may not be nearly as high over the long run as is expected. The purchased annuities would not be as large as is now anticipated.

14 United States of America

We come finally to the United States. Two major issues are present in its social security program (officially known as the Old-Age, Survivors, and Disability Insurance program). One is the controversy about the so-called retirement earnings test, under which persons who
are at least the normal retirement age (currently 65) but not yet age 70 receive reduced benefits when they have earnings that exceed a certain limit. If earnings are sufficiently high, all benefits are lost. The test is not applicable at age 70 and over, and it applies on a more stringent basis to beneficiaries under the normal retirement age. When benefits are received in later years, increases are given to reflect the benefits that are lost, but such increases currently are lower than those needed to provide actuarial equivalence. In 1993, such persons can earn up to $10,560 a year and still receive full benefits. But for every $3 of earnings above this limit $1 of benefits is lost.

A delayed-retirement credit (DRC) is given to individuals who lose benefits in this way, either because of not claiming benefits or because benefits are reduced thereby. For persons who reach age 65 in 1992 and 1993, the DRC is 4 percent per year of delay, pro-rated on a monthly basis. Under present law, the DRC gradually will be raised until it will reach 8 percent a year (which is about the actuarial equivalent) for persons who attain the normal retirement age in 2009 (then age 66). In other words, a person who then does not take benefits at the normal retirement age of 66 and waits until age 70 gets a 32 percent increase. This is about the same increase that a private insurance company would give under similar circumstances to a person who buys an annuity.

This test is unpopular with many persons. Critics say that it discourages persons from working and that, therefore, it is undesirable because work incentives are reduced. For many years, the author was a strong supporter of this test, under the simple but logical principle that retirement pensions should not be paid to persons who are not retired. After long deliberation about this matter and looking closely at the experience, it became evident that persons who had earnings of anywhere from about 50 percent to 150 percent of the average wage (currently, about $23,000 a year) receive little in their take-home pay if they continue working after age 65. Of course, for highly paid professionals who earn $100,000 or more per year, this is a different matter. But persons who earn $12,000 to $35,000 a year have great disincentives to work because the net additional income in their pockets from working is so small.

Therefore, it is clear that something should be done about this provision. The test should be eliminated for persons who are above the normal retirement age (currently 65), but under age 70, although they still should receive larger benefits if they continue working and do not collect benefits. They should receive 8 percent more per year in their eventual benefits under such circumstances. It is true that this
change would result in higher program costs, but only with respect to persons who attain the normal retirement age before 2008. (This is a low cost period for the program.) When measured over the long range, the average increase in cost is small (and can be met in several ways, none being especially painful—for example a temporary increase in the maximum taxable earnings base or, when changing the financing to a pay-as-you-go basis, not reducing the payroll-tax rates in the next two decades as much).

Next consider briefly the current financial situation of the U.S. Social Security program. Some say that it is going bankrupt, that it is in terrible financial condition, etc. Many in the United States think so, because they have heard or read about it somewhere. It is difficult to correct such misinformation. The program did have severe financial problems in the early 1980s, but these have been solved reasonably well.

At the end of 1992 the trust-fund balance was $331 billion; this is almost equal to one year’s benefit outgo. The trust fund is building rapidly, some $60 billion to $70 billion a year in the next few years and increasingly larger amounts for the next 15 years. From the short-range standpoint, the system is financially strong. But, as stated previously, one has to look beyond 20 years, because that is a short time in the life of a social insurance or pension program. Under the present method of financing, a large fund balance will be built over the next three decades according to the intermediate cost estimate in the 1993 Trustees Report. And it will reach a level of about $5 trillion in about 30 years. After then, however, it will decrease rapidly. In another ten years, it will be exhausted.

In the long run (after the year 2035), the system will have financial problems according to the current estimates under the intermediate assumptions. These can be solved at some time in the future, either by raising the contribution rates somewhat or by raising the normal retirement age (or both). There already has been a move in the latter direction. The normal retirement age slowly will increase under current law from the present age 65 (which has been in effect in the 56 years of operation of the system), beginning in 2003, to age 67 in 2027. An increase even to age 68 would have a significant financial effect.

The difficulty with the financing procedure for the U.S. system is that it is faulty in building a large fund and then drawing it down. Also—at least in this type of program and in the prevailing political process—building a large fund is undesirable. This may seem a strange thing for an actuary to say! Usually, if one is the actuary for
a pension plan, it would seem that the more money that one has, the better is the situation.

In this case, three good reasons exist why the procedure of building a large fund is undesirable. First, under the manner in which the federal budget is reported, the enormity of the deficit is hidden, to some extent, by the annual excesses of income over outgo of the trust funds. Second, the ready availability of these excesses for general purpose borrowing by the federal government could cause Congress and the executive branch to be less frugal than would be the case if borrowing were necessary from the private marketplace. Third, the mammoth size of the fund could cause irresistible pressures from the beneficiaries to overliberalize current benefits, thereby creating insupportable long-range costs.

Nonetheless, the amount of the present fund balance is needed as a contingency reserve in case an economic recession occurs. That balance (about one year's outgo) probably would get the system through any sort of business recession, even though income to the trust funds may be smaller than currently is projected.

The program has two trust funds, the Old-Age and Survivors Insurance Trust Fund and the Disability Insurance Trust Fund. These two funds usually are considered in combination when the financial status of the program is analyzed. The current estimates (intermediate) indicate that the combined funds will be exhausted in 2036, with the OASI fund lasting until 2044, but the DI fund only until 1995. This is not a significant problem, however, because the allocation of the OASDI tax rate can be changed slightly—as has been done several times in the past—to show both funds being exhausted at about the same time. Such reallocation would not have any effect on the taxes paid by workers and employers in the aggregate.

The foregoing discussion has not related to the Medicare program, which consists of hospital insurance (HI) and supplementary medical insurance (SMI). The former is financed by payroll taxes on almost the same persons as OASDI covers, while the latter is financed by premiums on the enrolled beneficiaries and by general revenues (which currently bear 75 percent of the cost). The HI program is estimated to have financial difficulties in the next ten years, its trust fund being exhausted in 1999 under the intermediate estimate. The SMI program rates are established in the law for years through 1995 at an apparently more than adequate level, and thereafter they can be adjusted by promulgations of the executive branch on the basis of experience.
Although the HI program has great financing problems over the long range, so too does the diverse health insurance system for the working population and its dependents. The solution to both sets of problems must be found simultaneously, perhaps by a radical change in the method of financing health care (which could mean the elimination of the Medicare program by the substitution of a universal system).

The solution to the foregoing problem of roller-coaster financing of the OASDI program is to change to responsible pay-as-you-go financing. Such a procedure was followed from the mid-1950s until the 1977 amendments. For more details on this matter, see Myers (1989).

The change to pay-as-you-go financing could be accomplished by lowering immediately the combined employer/employee tax rate by 1 percent for the next ten to fifteen years and then having the rate slowly increase over the years. Ultimately, the rate would have to be about 5 percent above the present 12.4 percent rate—just as would be necessary under present law after the trust funds are exhausted. As an alternative to such higher ultimate rates, the benefit costs could be reduced (e.g., by increasing the normal retirement age more than is provided for under present law). For more details on pay-as-you-go financing and its advantages, see Myers (1991).

References


On Becoming a Cost Effective Company

Robert D. Shapiro¹ and Barton H. Clennon²

Abstract

The 1990s financial services environment requires each life company to identify its distinct capabilities and competitive strengths and to build its future direction from these features. This demands a fundamental rethinking of traditional approaches to planning, organization, and financial management.

Key words and phrases: vision, quality, competitive advantages

1 Introduction

The realities of the 1990s operating environment have become painfully evident to life insurance company executives. The often-conflicting demands for strength, capital, service, and attractive prices are here to stay. Future survival requires the full complement of financial solidity, quality service, controlled expenses, and competitive prices. The increasing emphasis on full value requires that these factors be perceived favorably from the eyes of customers and agents and not merely reflect the hopes of management.

Because most life companies currently have expense levels that exceed pricing assumptions and have had expense excesses for many years, fundamental organizational structuring is necessary.

Whether such cost management and organizational restructuring initiatives destroy or enhance a company's value and long-term soundness depends largely on the strength of three factors:

¹ Robert D. Shapiro, FSA, CLU is president of The Shapiro Network, Inc. which was formed in 1987. Mr. Shapiro has been a consultant and investment banker to the insurance industry since 1965. He has extensive experience in strategic management, mergers and acquisitions, and financial management.

² Barton H. Clennon, FSA, MAAA is president of Clennon and Associates, Inc. Mr. Clennon was a partner in Milliman and Robertson, Inc., an international consulting actuarial firm. He has considerable experience in life insurance company financial reporting, appraisals, and mergers. His present interest is in life insurance company mergers and acquisitions.
Robert D. Shapiro and Barton H. Clennon  Cost Effective Companies

- Vision;
- Fit; and
- Commitment.

2 Articulating the Vision

Every successful company needs a clear vision, i.e., a picture of what it wants to become. Corporate vision must focus the company’s limited capital and human resources in a manner that exploits the company’s strengths and competitive potentials, emphasizes its commitment to quality and to success, and energizes its employees.

Vision is a “concrete description of where the company should be in five to ten years” (Shapiro, 1992). This definition emphasizes painting a picture, clear enough for all employees, agents, and competitors to see, of how the company will look after the company implements its plans. It provides a consistent framework within which all employees can make critical strategic and organizational decisions.

How will the vision shape a company’s cost management approach? First, it determines where and how the company must compete. This determines required price levels, which, in turn, drive allowable expenses. To compete and meet targeted profit goals, the company must shape its organization and related costs to live within the defined allowable expenses.

3 Fitting Company Capabilities

Each company needs to fit its skills and capabilities with the requirements of its vision and related product/service commitments. Different companies have the potential to excel at different things. For example, one company may emphasize innovative product/service features, while another may stress financial strength and safe, predictable returns. Each point of emphasis may demand different capabilities and management approaches, however, and may be valued by customers and agents in different ways. These differences in turn determine the specific level of expense that can be covered in prices.

If the vision is underpinned by special capabilities that provide distinct competitive advantages, the required value-added pricing (and related expense allowances) can be defined and tracked. To warrant the value-added pricing and costing required to reimburse the company for maintaining special capabilities, customers must appreciate these capabilities sufficiently to be willing to pay for them.

22
4 Commitment to Quality and Success

Although most insurers would argue that they have a strong commitment to quality and success, few have institutionalized this commitment. For example, quality efforts and rhetoric often are reflected in narrowly conceived projects with at best a temporary impact (much like many strategic planning retreats, mission definition sessions, and culture enhancement workshops!).

What does it mean to be committed to quality and success and to the related organizational focus needed to maximize marketing and service effectiveness and minimize costs? First, board level buy-in to the effort is essential to establish needed actions as long-term requirements that transcend current managerial personnel and employee agendas. Second, performance standards must be changed at all levels (i.e., corporate, unit, and individual employee levels). New performance standards (and consistently modified performance appraisal and compensation practices) send loud and clear messages that what is important has changed! Each employee ultimately will change what (and how) he or she does as it becomes clear what is important.

Commitment, like the definition of special capabilities, needs to be linked to vision. Without such linkage, there is no consistent framework for defining quality and managing the commitment to it. The result will be a shallow implementation program that will have limited impact and likely will disappear after a while.

When the vision linkage is present, the focus of commitment is to excel in the special capabilities that drive vision achievement. No company can be the best there is in all areas. Each has to establish its basis of competition and the core of its quality efforts based on the capabilities that can differentiate the firm in the marketplace.

5 Effective Expense Reduction

The narrow, meat-ax approach of many expense reduction actions taken by insurers in recent years has provided only painful, short-term fixes. Lost morale and paranoia may paralyze companies from doing the right things after the temporarily removed bloat returns.

The insurance literature is filled with comparisons of ratios of expenses to premium bases and ratios of expenses to the number of policies in force and unit costs. Although interesting and occasionally informative, these numbers provide life companies with little direction about the appropriate level of expenses and how to reduce costs to this level.
Horror stories of 10 percent to 20 percent across-the-board expense cut programs abound in the insurance industry. Because these expense cuts are generally arbitrary (and are perceived to be arbitrary), employee morale typically plunges and paralysis sets in as cuts are implemented. Productivity drops. Soon the company needs to fix some things to get the growth they seek. Staff is added, often in the same areas where it originally was cut. Costs rise, and the same competitive pressures that led to the original staff cuts reappear. A consultant is hired, expense cuts again are recommended ... and the cycle repeats.

Expenses cannot be analyzed in a vacuum. The framework for reducing expenses must encompass broader planning, financial, and pricing issues. The framework must be linked to the corporate vision to be understood and accepted by employees. Only within this broader framework can expenses be reduced without damaging the company and with an expectation that the reduced expense levels can be maintained.

The framework for effective expense reduction requires that the insurance company first:

• Establish clear, consistent profit and surplus objectives;
• Define target markets, distribution, and products/services;
• Identify future (not past) key competitors and how the company intends to compete against them; and
• Agree on the required level of price competitiveness and related sales expectations over the next four to five years.

Once this framework is clarified, the company’s allowable expenses can be defined and the organization can be reshaped to provide the required service within established expense standards. Normally a one to three year period is required to migrate from the existing expense infrastructure to a new one. The costs of organizational reengineering and related investments in new capabilities are key reasons why actual expenses will run in excess of allowable expenses for a period of time.

6 An Expense Model

Every company must balance what it spends against the price customers/agents are willing to pay for the company’s products and services. An insurance company’s expenses can be considered in this way:
The shaded areas represent expenses that must be eliminated by the company for expenses to cover prices. Stated another way, an insurance company only can afford to spend on extra services what customers and agents are willing to pay for such services. There are some costs such as regulatory compliance, of course, that may not be appreciated by customers and agents but that cannot be eliminated. These costs generally would be mandated for all insurance companies and hence would be reflected in expenses of a low cost insurance company.

Although this broad analysis may seem obvious, it is difficult to identify and cost the specific value-added services that are provided by the company. Insurance companies typically analyze costs at department or function levels and not by process or task. Further, most companies continue to define what they do without substantial analysis of what customers and agents really want and need. Hence, the data needed to determine and eliminate the cost of unappreciated extra services provided by the company are not available.

Nonetheless, each company needs to push its way through this exercise as best it can. While this value-added quantification and analysis is being developed, those existing tasks and operations that are ineffective or duplicative can be eliminated. Hence, the expense reduction process can achieve some immediate successes while the foundation for more fundamental reductions is being established.

### 7 Refocusing the Organization

It is difficult, if not impossible, to reach full potential in providing valued service without a clear vision of what the company wants to be. A coherent vision will articulate target markets, expected key competitors, required financial standards (e.g., price levels and allowable expenses), and desired operating approaches. The vision also will define activities that must be accomplished in order for the company to be successful.

An insurance company’s organizational structure must be reshaped to focus on these critical activities. New priorities and new employee by employee daily agendas must be established. Activities that are not consistent with the vision must be eliminated.
For example, let’s say the vision demands that the company be the leader in serving the insurance needs of small employers. Assume further that the company historically has organized its operations in life, annuity, health, and group segments. The vision undoubtedly places great value on activities that will capture and serve small employers. It likely devalues other activities that fit the historical operation segmentation (such as stockbroker sales promotions within the annuity line) but that are not within the concentrated focus of the small employer vision. Similarly, many activities that historically were developed within one or more of the life, annuity, health, or group segments need to be reconceived and streamlined within the new vision.

Reshaping operations to provide targeted, quality products and services demands clear articulation of where the company must have the highest quality to achieve the vision. Each current activity should be analyzed by asking the following questions:

- Is the activity essential to the role of the unit in which it is being performed? Are the unit and the activity vision consistent?
- Could the frequency, scope, or precision of the activity be reduced without a significant negative impact on the ability to manage or operate?
- Could the company eliminate, simplify or move all or part of the activity?
- Could the company reduce the cost of performing the activity?
- Could the company improve control methods?
- Are functions and activities grouped in the most effective way, given the stated vision?
- Are jobs designed for efficient performance of assigned tasks?
- Who pays for the activity, e.g., existing policyholders, future policyholders, surplus, etc.?

8 An Example

Let’s take a simple example. Assume we have a life insurance company (ABC Life) that writes only $100,000 face amount life insurance policies, each with an annual premium of $1,000. Assume further that ABC Life has 100,000 policies in force, writes 20,000 new policies per year, incurs noncommission expenses of $20 million per year, prices for noncommission expenses using $500 per policy first year and $50 per policy in renewal years, and has clear profit and surplus goals.
The annual expenses allowed by current prices in the current year amount to $15,000,000 (i.e., 100,000 \times $50 + 20,000 \times $500). Hence, ABC Life's noncommission expenses are running 133 percent (i.e., $20,000,000 + $15,000,000) of the expenses allowed for in its pricing ... or $5,000,000 per year of excess expenses.

What can ABC Life do? Three of its options are:

1. Increase prices to allow for 33 percent more in expense. For example, the prices could be increased to allow for $668 per policy first year and $67 in renewal years. Issues include:
   - Will prices still be competitive enough to write 20,000 new policies each year?
   - Can existing policyholders be charged $67/policy? If so, will they continue their policies if their price is increased?

2. Cut 25 percent of expenses (from $20 million per year to $15 million per year). Issues include:
   - How and where should the expenses be cut?
   - Can servicing and support activities be maintained at a high enough level to keep customers and agents happy (and persistent)?

3. Sell more business. If ABC Life can get to where it writes 27,000 policies a year and has 135,000 policies in force, its current pricing expense allowance would provide over $20 million per year (133,000 \times $50 + 27,000 \times $500).

How should ABC Life proceed? Many companies make a judgment to take one (or a combination) of these options, relying more on hope than solid action plans that customers will continue to buy, expenses can be cut, and/or more sales can be developed. The demands of today's complex marketplace will require proactive, well-planned actions even in companies that have a track record of success.

History tells us that hope rarely brings success. During the 1980s the majority of life companies had general expenses that were in excess of those allowed in their pricing. Many of these companies embarked on one or more plans to bring expenses in line with allowables, yet most of these plans failed to meet expectations. The main reason for failure was that extrapolated growth and/or cost reduction projections were not realistic in an environment characterized by intense competition, increasing capital pressure, and proliferating regulatory and administrative demands.

Given the above options, the place for ABC Life to begin is at the intersection of ABC Life corporate potential with the opportunities and realities of the expected future financial services marketplace. Where can or should ABC Life compete? Who will its com-
petitors be? What price levels will be required in this market? How much money will these required prices allow ABC Life to spend? To answer this last question, ABC Life must make a judgment about the anticipated relationship between price (and related expense allowables) and sales levels to determine specific price and sales objectives.

Once the target price and sales levels are established, the allowable expenses are determined. Let's hypothesize that ABC Life's analysis determines that:

- Its vision requires it to keep its prices at current levels. Hence, current expense allowances need to be maintained;
- By refocusing its marketing and product approaches to be consistent with its vision, it can write 25,000 new policies per year (with inforce stabilizing at 125,000 policies per year in five years); and
- Current expenses would be reduced from $20,000,000 per year to $16,100,000 per year if the organization were reshaped to eliminate inefficiencies and support modified operations more effectively. It will take three years to reshape the organization, however.

A (simplified) five year quantification of ABC Life's new future expectations might look like this:

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>ABC Life's Five Year Future Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>Expected Policies New</td>
</tr>
<tr>
<td>Year</td>
<td>Policies</td>
</tr>
<tr>
<td>1</td>
<td>25,000</td>
</tr>
<tr>
<td>2</td>
<td>25,000</td>
</tr>
<tr>
<td>3</td>
<td>25,000</td>
</tr>
<tr>
<td>4</td>
<td>25,000</td>
</tr>
<tr>
<td>5</td>
<td>25,000</td>
</tr>
</tbody>
</table>

*A simplified approximation reflecting potential basic cost reductions, re-engineering costs, and inflation.

ABC Life expects to bring its expenses in line with prices and sales within three years. Current inflation expectations will create another expense deficiency after five years, however, unless something else is adjusted. ABC Life may have to wait a year or two to
reexamine the results of its restructuring and changes in the competitive environment for additional ideas.

9 Conclusion

The 1990s require that each life company look deep into its corporate soul to find its distinct key to future success. Shallow pricing and sales gimmicks will not work and can undermine or destroy a company quickly.

Few companies can afford to dissipate value (which is equivalent to capital). Unless a direction can be sculpted that is anchored in a company's special capabilities and aligned with the realities of the competitive marketplace, the company runs an unacceptable risk of self-destruction. For most companies, a fundamental rethinking and an approach to planning, organization, and financial management similar to that described in this article are necessary for long-term financial health and viability.

References


Robert D. Shapiro
The Shapiro Network Inc.
312 East Wisconsin Avenue, Suite 700
Milwaukee WI 53202

Barton H. Clennon
1112 Appleland Drive
Wenatchee, WA 98801
The Process of Pension Forecasting

Michael Sze*

Abstract

This paper explains the process of pension forecasting. It discusses the common purposes and uses of pension forecasts, the major steps involved, and the principal limitations of these forecasts.

Some insights into each stage of the forecasting process are provided. Among the stages discussed are: the background research to be performed; the selection of scenario assumptions; shortcuts used in the actual performance of the forecast; review of the forecast results; and communication of the forecast findings.

Key words and phrases: projection, simulation, stochastic modeling, scenario

1 Introduction

Funding retirement obligations has become a significant part of corporate financing. It is not unusual for a plan providing rich retirement benefits with indexation or one with substantial unfunded past service liability to require an annual contribution in excess of 15 percent of payroll. The unfunded liabilities of some companies' pension programs are equal to a sizable portion of their net worth. Union negotiation settlements hinge more and more on pension agreements. As a result, many companies include a pension forecast1 in their regular financial planning process.

The responsibility for providing such a pension forecast typically is delegated to the actuary. Most actuaries are familiar with the

* Michael Sze is a Fellow of both the Society of Actuaries and the Canadian Institute of Actuaries. He received his Ph.D. degree in mathematics from the Ohio State University and currently is a partner of Hewitt Associates. He is the chair of the Society of Actuaries Retirement Systems Research Committee, as well as a member of the Canadian Institute of Actuaries Investment Practice Committee. While the author takes full responsibility for any errors in this article, he would like to acknowledge, with gratitude, the valuable comments provided by Ms. Rita Lawlor, Ms. Milena Francia, Ms. Megan Duke, and Mrs. Elsie Sze in the preparation of this article, as well as many helpful suggestions by the referees.

1 The terms projection and forecast are used interchangeably in this paper and in the pension actuarial literature in general.
basic mathematics involved in a pension forecast because of their training and education. Many inexperienced actuaries, however, are not familiar with the actual process of a pension forecast. In fact, some pension valuation actuaries actually have difficulties making forecasts, primarily due to the fact that they do not fully recognize the difference in emphasis between an actuarial valuation and a forecast. The former focuses on the present; its aim is to provide an accurate assessment of the funded status and cost of a pension plan under the current legal and accounting environments. The latter is directed toward the future trend of pension costs under varying economic or demographic scenarios.

The consequence of not understanding the forecasting process fully can be costly. At best, the actuary may have difficulty explaining the cause and effect of some economic variables. At worst, faulty assumptions or logic can lead to erroneous conclusions with detrimental effects to the company. Because there are many variables involved in the process, there is a real danger that errors often are not detected until the damage has been done.

The purpose of this paper is to share some of my experiences in pension forecasting, to provide some insights regarding the process, and to point out some possible pitfalls. Because of the complex nature of a pension forecast, it is impossible to cover every possible situation. This article, however, can be used to assist in more diligent planning of each forecast; it is not a cookbook to be followed in every step of the process. Readers are assumed to be familiar with the basic techniques and mathematics of the projection process. These assumptions represent management's best guess of future economic or demographic scenarios.

This article is organized into six sections, each of which is briefly described below.

- **Preparation for a Forecast**: This section discusses the major considerations and background research that must be performed before embarking on the forecast. Most problems confronted in pension projection originate from insufficient preparation;
- **Choice of Scenario Assumptions**: This section covers some basic considerations underlying the choice of scenario assumptions. These assumptions represent management's best guess of future economic events. Such assumptions control the projected results and must reflect the principal objective of the projection;
- **Performing the Forecast**: This section discusses the choice of the projection method. The purpose and needs of the sponsor determine the scope of the forecast;

---

2 Readers interested in the details of the pension forecasting process may refer to Lorisz (1993), Sze (1997), or Schnitzer (1977).
2 Preparation for a Forecast

The importance of preparation cannot be overemphasized. Even the most experienced actuary must have on hand a detailed preparation of what he or she plans to accomplish with the forecast. A detailed preparation should consider: (1) the purpose of the forecast; (2) the sponsor, the industry, and the economic environment; (3) the demographics of the population; (4) the pension plan, the valuation methods, and the actuarial assumptions; and (5) past plan experience and the funded status of the plan.

2.1 Purpose of the Forecast

Unlike funding and expensing valuations which are required by governmental regulations, there are no legal or accounting rules requiring pension projections. The request to perform a pension projection study usually originates from plan sponsors who need answers to specific questions concerning their pension plans. Before the actuary begins the study, it is important that he or she knows what those questions are and the reasons for the questions. Knowing the purpose of the forecast will lead to a better understanding of the sponsor's funding and expensing expectations and the sponsor's risk tolerance. An integral part of the forecast is the testing of the achievability of the sponsor's objectives under legal, accounting, and economic constraints. Understanding the sponsor's expectations and risk tolerance also will provide guidance on the choice of scenario assumptions, the scope of the study, and the best way to communicate the forecast's findings.

The emphasis of a forecast depends a great deal on its purpose. A forecast that is part of the regular corporate financial planning process may have as its goal one of the following: (1) to determine the stability of pension contributions and expenses; (2) to devise funding
and investment strategies that will minimize unexpected fluctuations in contributions and expenses; or (3) to devise an asset/liability matching strategy to minimize undesirable deterioration in the funded status of the plan.

A few examples may illustrate some of the considerations involved:

- **Downsizing**: In a downsizing operation, the forecast should anticipate significant aging of the group, the possibility of an employer-initiated early retirement program, and a decrease in population size. The chance of plan terminations typically cannot be ruled out. There is a need to monitor the risk of having to fund the entire plan deficiency over a short period of time. The alternatives that should be considered are amortization and bond immunization. In one such study, a sponsor had to consider the impact of the timing of plan termination after a downsizing process. Figure 1 shows the funding impact of plan termination in different years, assuming that plan termination deficiency is amortized over five years. It further demonstrates that the funding pattern is practically the same (except the incidence of payments), irrespective of the timing of the plan termination decision. The actuary in this case was instructed to monitor interest rates for the sponsor. An annuity contract was placed at an opportune time which allowed the sponsor to save millions of dollars on the plan termination cost;

- **Changing Employment Pattern**: As a result of the demographic pattern of aging shown in the United States and Canada, many retail companies have experienced a significant shift in hiring patterns. Companies often want to know the impact of such demographic changes on future pension costs. In such a study, the emphasis must be to balance the need for adequate retirement benefits for the employees with the need for stable pension contributions and expense for the employer. The alternatives that should be considered are plan design changes (such as a change from a defined benefit plan to a defined contribution plan) and funding basis changes (such as changes in retirement age and turnover assumptions). The scenario assumptions used must reflect age and sex distributions of new employees as well as more realistic pay and termination patterns for these employees.

    In the early 1980s, a major department store expected that new hires would be substantially older and would include a larger percentage of females. Many of the new hires would be the secondary wage earner of the family and might not be as career-aggressive as were previous employees. A forecast study was commissioned to study the pension cost impact of these demographic changes as well as to suggest alternative plan designs. The plan had a sizable funding surplus, so the contribu-

---

3 Readers interested in the theory and application of bond immunization should see Redington (1952), Tilley (1980), and Bader (1983).
Figure 1
XYZ Corporation Retirement Plan Impact of Shutdown

[Diagram showing the impact of shutdown on required contributions over years 1992 to 2002.]
ion pattern was not a major concern. The cost considerations were directed to the trend of pension expense as a percentage of payroll. The study analyzed the net cost increase after taking into account the offsetting impact of aging, a more moderate rate of salary increases, and a higher turnover rate. The expense increase was moderate and was considered to be manageable by the plan sponsor. The defined contribution alternative, while helpful in stabilizing pension cost, was considered to be too drastic and was deemed to provide unsatisfactory retirement income for employees. In the end, no major plan design changes were made. There were, however, some changes in actuarial valuation assumptions to reflect more realistic expectations of salary progression and turnover pattern;

- **Financial Planning to Stabilize Pension Expenses:** Financial Accounting Standard No. 87 requires that the discount rate used to determine pension liabilities and service cost must be based on the current market interest rate. Plan sponsors feel vulnerable to unpredictable shifts in economic situations, especially given the volatility of market interest rates in recent years. Also, the fluctuating investment returns of pension funds add to the uncertainty of the pension cost. A forecasting study may be ordered to determine a stable projected pension expense trend. The alternatives considered typically include asset/liability matching. Numerous other articles have covered asset/liability matching and immunization.4

Many forecasts have been prompted by investment advisors. The actuary is asked to provide the liability and cash flow trends of the pension fund. A forecast is performed to test investment policies against the deterministic liability trend in order to find the investment mix that best protects the surplus of the plan. These forecasts often result in a recommendation for a higher investment concentration in bonds.

This approach to projection misses the interplay between assets and liabilities. A detailed stochastic projection involving both assets and liabilities (usually referred to as asset/liability modeling) will tend to produce substantially different results. For example, an inflationary environment will impact both wage increases and investment returns simultaneously. Only an asset/liability modeling process will be able to capture the correlated events between assets and liabilities; see Beekman (1980), Redington (1952), and Tilley (1980); and

- **Postretirement Medical Benefits Forecast:** Many companies are interested in investigating the immediate and continuing impact of FAS 106 rules. These rules require companies to book liabilities and expenses for postretirement medical and other benefits. Because of the scarcity of background information, many attempts

---

4 A discussion of asset/liability matching and immunization is beyond the scope of this article. For more information on this topic, see Beekman (1980), Tilley (1980), and Redington (1952). For more on immunization and how it may help to stabilize pension cost, see Daskais and LeSueur (1983) and Sze (1993).
to forecast the impact of FAS 106 have been performed that treat the postretirement medical benefit payments as a stream of escalating annuity payments during the life of the beneficiaries, with the escalation reflecting medical inflation. The present value of these payments usually is determined by using expected investment returns and the mortality and termination decrements used in pension valuations. Other considerations usually include alternative expensing bases and benefit designs.\(^5\)

There are two potential flaws to forecasts performed in the manner described in the last paragraph. First, the pattern of postretirement medical benefit payments is different from that of an escalating annuity. A major portion of medical expenses are incurred during the last few years of a person's life; see Riley and Lubitz (1989).\(^6\) Second, the present value calculations in many FAS 106 projections are based on mortality and termination rates used in pension valuations. Mortality rates used in a pension valuation often overstate actual experience, while termination rates typically understate actual experience. Such discrepancies may have a significant impact on the liability and service costs calculated.\(^7\)

Aside from the flaws in many FAS 106 studies, the forecast results still may present valuable information to plan sponsors. After the initial shock of the drastic cost impact of providing these benefits, many plan sponsors would explore other plan design alternatives such as requiring employee contributions, establishing maximum benefit limits, or replacing welfare benefits by additional pension benefits.

Advance funding of this obligation may be considered. Funding alternatives often investigated include funding through the pension plan based on Internal Revenue Code (IRC) Section 401(h) or establishing a separate trust under IRC Section 501(c)9. Under some restrictive conditions, IRC Section 401(h) allows funding of such postretirement health benefits in a pension plan. IRC Section 501(c)9 allows prefunding of welfare benefits under limited conditions; see Hess, Becker, and Snyder (1991) and Kra and Resse (1992). Expensing alternatives include immediate recognition of past service liability or amortizing this liability over the expected future service of the employees.

The above examples illustrate the need for the forecast to reflect the purpose of study. It is important to note that because each project is initiated to address a specific problem, the actuary should provide

---

5 For more on funding postretirement medical benefits, see Roccas, Sobel, and Ullman (1990) and Veach, Cotter, and Meyers (1992).

6 Further research is needed to determine the actual pattern of payments and the impact of the proper cost attribution. Studies in these areas are currently being undertaken by the Society of Actuaries.

7 See Vaughн (1992) for more on realistic termination experience.
not only an explanation of the cause and effect of the issue in question, but also alternatives for solving the problem. It is this last requirement that makes a forecasting study more challenging to the actuary and valuable to the sponsor than a regular funding or expensing valuation.

2.2 Sponsor, Industry, and Economic Environment

Forecasting studies never are performed in a vacuum. They are performed in the context of other economic events. A trend of escalating pension costs may be tolerable for a utility company. The rates that a utility company charges its consumers typically are fixed on a cost-plus-margin basis. Thus, any increase in operating cost is passed to the consumers. On the other hand, the same cost trend may be detrimental to a manufacturing company undergoing severe downsizing in a recessionary economic environment. In such an economic climate, the revenue is limited by price competition. Severe downsizing, however, typically entails sizable escalation in pension cost.

Before beginning a projection study of a pension plan, it is important to understand the financial strength of the plan sponsor, as well as the significance of the pension cost in the operating budget of the company. A company with ample resources may be able to tolerate more fluctuation in the pension cost, so the funding time horizon may be longer. Thus, the goal may be to achieve the most favorable long-term financial results, even if it means taking more risks in the interim. On the other hand, for a company with limited resources or whose pension cost is a significant portion of its total budget, care must be taken to ensure acceptability at each forecast year. An unexpectedly high cost at any point may be unacceptable to the company, requiring immediate management attention, which often results in funding and/or investment changes. The constraints for such a forecast are much tighter, and results for each forecast year must be examined carefully.

It is important to understand the business of the plan sponsor. This often dictates the hiring, promotion, and termination patterns of the company. Knowledge of the industry in which the plan sponsor operates provides insights into the growth or retrenchment pattern of the overall employee population, as well as the volatility of such a pattern. Such knowledge determines the choice of demographic scenario assumptions.

Many forecasts are commissioned when the sponsor has a problem that needs addressing. Often these forecasts are performed in times of economic downturn. The future economic outlook is critical in assessing

Michael Sze Pension Forecasting
a plan sponsor’s tolerance for cost fluctuations. For a plan sponsor with a severe cash flow constraint, it is of paramount importance that the forecast addresses both the current economic outlook and the consequence of further economic downturn.

2.3 Demographics of the Population

The demographics of the employee population determine not only the current year’s cost of the pension plan, they also dictate the future retirement and termination patterns of the plan. Where the cash flow forecast is critical, a careful study of the demographics of the current employee population is vital. Furthermore, a less mature employee population does not have as much pending pension obligation as a more mature population and may have greater tolerance for economic fluctuations. A careful study of the population demographics provides much insight into the trend of the future costs of the plan.

2.4 Pension Plan Valuation Methods and Actuarial Assumptions

The impact of economic factors on the future pension cost depends on the plan’s valuation methods and actuarial assumptions. Thus, it is important to review these valuation bases before embarking on the forecasting process. For example, the company’s contribution, expressed as a percentage of salary for a defined contribution plan, is insensitive to salary changes. The pension cost of a final average salary defined benefit plan, however, is affected greatly by salary experience, especially if the pension plan benefit is integrated with Social Security. The pension cost of a career average salary defined benefit plan is less volatile with respect to salary experience.

Pension costs under aggregate cost methods are typically less sensitive to the effect of aging populations than are pension costs under individual cost methods. The entry age cost method (among individual cost methods) tends to provide a more stable cost pattern with respect to an aging population than does the unit credit cost method. Unit credit normal cost represents the present value of benefits earned during the valuation year. As the population ages, the normal cost escalates. Entry age normal cost represents the average of such nor-

---
8 For a detailed analysis of pension costs methods, see the texts by Anderson (1990) and Berin (1989).
mal costs over the career of the participant. It is more stable relative to the aging process of the population.

A more aggressive actuarial valuation interest assumption anticipates higher investment returns and thus provides less opportunity for asset gains. A higher valuation salary scale assumption anticipates higher cost increases due to pay increases and, therefore, provides greater opportunity for pay gains.

2.5 Past Experience and Funded Status of the Pension Plan

Some economic variables are difficult to predict because their behavior is independent of the past. Many pension plan variables (such as turnover and promotion patterns), however, are not independent of past experience and can be projected with a certain degree of accuracy. A study of past experience of these variables thus provides valuable information for the future. Overall, ignoring past experience in a forecast study is likely to lead to worthless results.

Temporary investment and other experience setbacks may be tolerable for plans that have huge funding surpluses. The experience impact on a plan’s funding requirement can be drastic for plans that are only marginally overfunded; therefore, pension forecasts must recognize the funded status in the selection of scenario assumptions.

3 Choice of Scenario Assumptions

Because scenario assumptions control the occurrence of certain key economic events that may impact future pension cost, the proper choice of assumptions is vital to the usefulness of the forecast. These assumptions must echo the purpose of the study, recognize both the plan’s and the sponsor’s characteristics, and reflect past experience of the plan.

The choice of scenario assumptions must be a joint effort between the actuary and the plan sponsor. The plan sponsor’s input is critical because scenario assumptions should reflect management’s best estimate of future economic events. Furthermore, the sponsor has the best understanding of the needs of the company, the financial risks that it can tolerate, and the company’s objectives. The sponsor may not have analyzed past experience as carefully as the actuary, however, and may not have ready access to economic and investment data or have as much understanding of the implications of the choice of some assumptions as does the actuary. Furthermore, the bias of the plan sponsor, whether intentional or not, may prejudice the objectivity of
the analysis. Thus, it is the responsibility of the actuary to provide guidance on the cause and effect of the choices. Where there are doubts about some selected scenario assumptions, alternative assumptions should be tested.

During the process of choosing scenario assumptions, an often asked question is: "What is the valuation assumption?" Such a question usually reflects a lack of understanding of the basic purpose of these two types of assumptions. It is the actuary's responsibility to explain the difference between forecast scenario assumptions and actuarial valuation assumptions. Actuarial valuation assumptions typically contain a margin of conservation that should be removed in the choice of scenario assumptions for forecasting. For instance, the commonly used valuation mortality table (e.g., 1983 Group Annuity Mortality Table) provides mortality rates that are 10 percent lower than the underlying experience obtained by mortality studies of the population over the same period; see Committee on Annuities (1983 and 1987). Similarly, typical withdrawal tables provide turnover rates that are lower than actual experience; see Vaughn (1992). These subtle differences often are not explained clearly to the plan sponsor. As a result, valuation turnover assumptions often are chosen by default to be the scenario assumptions. For pension plans where the death benefit is comparable to the projected retirement benefit, using a valuation mortality assumption for the scenario mortality rate may not distort future pension cost greatly. Where death benefits are payable in a lump sum, the cash flow pattern will be understated if the actual number of deaths exceeds the expected number of deaths. The distortions introduced by conservative turnover assumptions, however, may be even more significant, as the turnover rate is typically much higher than the mortality rate.

The set of scenario assumptions should include the following groups of assumptions: demographic, economic, and simulation assumptions. This article will not provide a detailed explanation of each scenario assumption. (Interested readers should see Sze (1987) for details.) We will provide, however, a few critical comments on some of them.

3.1 Demographic Assumptions

Demographic assumptions are used to project future employee populations. Such assumptions include the mortality, disability, ter-

---

9 See Lorisz (1993) and Sze (1987) for more detailed discussions.
mination and retirement patterns and the number and distribution of new entrants with respect to sex, age, and pay.

Usually mortality and disability scenario assumptions are only age specific. The termination assumption, however, should vary by age and duration and should show a higher turnover pattern during the earlier years of the employees' careers. The retirement pattern should be distributed over the eligible retirement ages. Although the number of new entrants may differ from year to year, the distribution by sex, age, and pay usually is assumed to be the same during the projection period; see Jackson, Haley, and Wendt (1989) and Sze (1987).

For a small pension plan, a significant demographic change would produce a major impact on the trend of pension costs. The assumption of such demographic changes usually is specified by the sponsor.

3.2 Economic Assumptions

Economic assumptions are used to project and determine the assets and liabilities of the plan during the forecast period. These assumptions include: an inflation rate; real or nominal investment rate of return; a salary increase; flat dollar benefit rate increases; and government benefit increases.

Actuaries traditionally assume that the real investment returns and the real rate of salary increases are constant throughout the forecast period. Thus, nominal returns on assets and projected pay increases only fluctuate with inflation. In addition, investment returns and salary increases always move in the same direction. Salary losses consequently are compensated by investment gains and vice versa. In the end, the projected pension cost is more stable than may be expected. In reality, nominal investment returns often are correlated negatively with inflation. (See Table 1.) In times of high inflation, real salary increases may be close to zero. Under such circumstances, pay losses resulting from high inflation rates may be coupled with substantial investment losses. It would be imprudent for actuaries to ignore this worst case scenario.

3.3 Simulation Assumptions

Simulation assumptions are needed to perform stochastic asset/liability simulations. They typically include the economic assumptions discussed above; the real rate of return and the standard deviation for each asset class; the real salary and real benefit
increases and their standard deviations; and the correlation between each pair of variables, as well as the correlation of each variable with inflation.

Many of these assumptions, especially the correlation factors, often are chosen arbitrarily, mainly because both the actuary and the plan sponsor may not have a good feel for the significance of these assumptions. Improper choice of assumptions, however, may distort and invalidate forecast results. Actuaries who wish to develop their expertise in asset/liability simulations are advised to test alternative assumptions to build their intuition in this area.

The following is a correlation matrix of inflation and the real returns of some common asset classes in the United States from 1926 to 1988:

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>T-Bill</th>
<th>LT Bonds</th>
<th>S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1.00</td>
<td>-0.72</td>
<td>-0.55</td>
<td>-0.24</td>
</tr>
<tr>
<td>T-Bill</td>
<td>-0.72</td>
<td>1.00</td>
<td>0.57</td>
<td>0.14</td>
</tr>
<tr>
<td>LT Bond</td>
<td>-0.55</td>
<td>0.57</td>
<td>1.00</td>
<td>0.22</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>-0.24</td>
<td>0.14</td>
<td>0.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sources:
- Three Month U.S. Treasury Bill Yield
  1926-1941 Homer Sydney. *A History of Interest Rates: 2,000 B.C. to Present*, Table 51, Part II
  1984 on *Federal Reserve Bulletin*, Table 1.35, Interest Rates, Line 18
- Long-Term U.S. Government Bond Return
  1954-1977 20 Year Treasury Constant Maturity Yield Percent, Average of Daily Figures, *Federal Reserve Bulletin*, Table 1.35, Interest Rates
  1986 on 30 Year Treasury Constant Maturity Yield Percent, Average of Daily Figures, *Federal Reserve Bulletin*, Table 1.35, Interest Rates
- Standard & Poor's Composite Return
  Standard & Poor's Statistics Service: *Security Price Index Record*
4 Performing the Forecast

A critical decision in performing a forecast study is the choice of forecasting method. The reader is assumed to be familiar with the mechanics of the following common forecasting methods:

• The simplified forecast (also called the back of the envelop approach);
• The seriatim forecast;
• The forecast based on group data; and
• The stochastic asset/liability simulation.

Instead of reviewing the details of each method, we will provide some hints on the choice of the method.

Over the years I found that almost three quarters of all projections may be performed using the simplified approach. The simplified approach is quick and easy to do and provides reasonable results that reflect the intuition of the actuary. Because of repeated iterations involved in this process, however, inherent estimation errors escalate geometrically. For instance, a 5 percent overestimation of liability each year will compound to over a 60 percent error in ten years. These projection results are typically not reliable after the first five to ten years, depending on the experience of the actuary. Furthermore, the method does not capture the impact of demographic changes readily. Nor is it capable of ascertaining the subtle effect of the application of individual benefit limits. Finally, because this approach is based on the intuition of the actuary, the results must be reviewed carefully by an experienced actuary. The real danger lies in the fact that an inexperienced actuary may produce misleading results without realizing the mistake.

Where detailed results are required, a seriatim or group data forecast is recommended. Because a detailed seriatim forecast is costly, some data grouping usually is deemed necessary. Grouping into age/service/pay cells typically is satisfactory. Highly paid employees and employees close to retirement should be identified separately, however. The termination/retirement patterns for these groups of employees must be handled separately because of their potential impact on pension cost and cash flow of the plan.

10 It is not the intention of this article to provide a detailed explanation of the various projection methods or the background mathematics. Interested readers are encouraged to study Schnitzer (1977); Jackson, Haley, and Wendt (1989); Lorisz (1993); and Sze (1987).
Asset/liability simulation is used mostly in the process of establishing an investment policy. This type of simulation is very good for analyzing the risk factors involved in a funding or expensing policy. Through numerous asset/liability simulations, one can test a policy under different economic situations. Based on the simulated outcome, the sponsor better may understand the down side of the policy decision. A major difficulty in this type of forecast, though, is in establishing the input assumptions. The actuary should try different sets of input assumptions to gain insight into the effects of the different choices. Another difficulty with these forecasts is the volume of output information generated. The actuary should study the outcome generated carefully and distill these results to the bare essentials before attempting to provide meaningful communication.

5 Review of Forecast Results

The review is the most important technical step of the forecasting process. As mentioned above, the voluminous output generated by this process requires that the actuary diligently sort the results to make sure they make sense and that they address the questions asked.

An important criterion to bear in mind in the review process is simplicity.

Something must have been wrong if there are no simple explanations for the forecast results. Probably some important factors have been overlooked or have been included improperly in the forecast.

A useful tool to check for reasonableness is to perform a projection of pension liabilities and cost using a simplified projection performed under the same scenario assumptions.

A careful review of the simplified projection’s results typically will reveal details that have been overlooked or some alternative perspective that warrants further considerations. The full projection then must be revised to reflect these requirements. This cycle of forecast, review, and refinement usually is repeated several times until the actuary is satisfied that all results make sense and the different perspectives have been analyzed.

To date, there is no completely objective criterion for judging the validity of the forecast results. The following are some helpful hints on checking the internal consistency of forecast results:
Under each funding or expensing basis, the ratio between any pair of the following items is usually rather stable: valuation payroll, normal cost, present value of future compensation, and present value of future normal cost. There is a slightly less stable relationship between each pair of the following items: active accrued liability, active accrued benefit value, and active vested benefit value;

The relationship between the normal cost and the accrued liability usually shows a stable trend, reflecting gradual increases or decreases in the average age and service of the group;

For a mature population, the ratio of the inactive accrued liability to the active accrued liability is usually quite stable. On the other hand, for an immature population, that ratio tends to increase over the projection period;

When comparing results under different funding or expensing bases, note the following relationships:

- The ratios of corresponding items under the different bases should remain stable;
- Normal cost increases are more sensitive to the aging pattern of the population under the unit credit cost method than under the entry age normal cost method; and
- For a final average pay plan, the increases in the accrued benefit value reflect the total pay increase while the increase in the accrued liability only reflects the actual pay increase in excess of the salary scale assumption.

6 Communication of Forecast Findings

From the plan sponsor’s perspective, communicating the finding may be the most critical step of the entire process. The actuary must be careful not to confuse the sponsor with the endless stream of numbers from a forecast report. It is important for the actuary to understand forecast results through the review process and essential that he or she be able to share this understanding with the client.

The actuary may believe at the end of a project that the conclusions of the study are self evident. But the forecast findings become obvious to the actuary only as a result of weeks of work and self-education. The final challenge is to educate the audience in the course of a one or two hour meeting.

A useful suggestion is to stay focused on the initial questions asked. Even though millions of numbers are produced, only those relevant to the purpose of the projection should be presented. The fewer the details shown, the more the concept will be absorbed by the listener.
Many actuaries experienced in pension forecasting have told stories of how they were trapped years after a forecast into explaining why their forecast results differed from actual valuation results. It is easy to blame the sponsor’s ignorance of the estimations involved in the forecast process. Knowing the limited precision of the results, however, we question why such details ever were communicated in the first place. Were the actuaries unaware of the imprecision involved? Were the actuaries trying to attribute too much exactness to the process?

In spite of the high volume of output data, the principal purpose of the forecast is to analyze trends under various scenarios. Both the trends and the comparisons are easiest to visualize through the use of graphs. Forecasters should experiment with different ways to graphically present their results.

7 Other Considerations

Although both pension actuarial valuations and forecasts are based on the same mathematical principles, the uses of their calculations are quite different. The purpose of an actuarial valuation is to establish the funding and expensing requirements for the year. It is performed under regulatory or accounting rules. The basis of an actuarial valuation tends to be conservative. The results provided often are considered to be exact. On the other hand, the purpose of a pension forecast is to test the future cost impact of some expected or proposed changes. The emphasis is on the future trend of the cost. The important result is the cost comparison under different scenarios. This difference in the basic purpose of the two processes is reflected in several factors:

- **Assumptions**: Valuation assumptions have margins of conservatism. Forecast scenario assumptions tend to be realistic;
- **Results**: Valuation results often are used to derive exact funding and expensing requirements. Forecast results should be shown as estimates;
- **Time Horizon**: Valuation results are only applicable to the current year. Forecast results may cover ten or more years;
- **Alternatives**: Valuations provide pension cost under specific sets of conditions. Pension forecasting usually is performed to compare pension costs under several alternatives. The goal is to choose the alternative that best reflects the objectives of the sponsor;
- **Variation**: Because of the extended outlook and additional alternatives considered in a forecast, there tend to be more variations
in this process. The forecast is certainly more challenging and more interesting (to me, at least) than a valuation;

• **Estimates:** Forecasting has its limitations. It is important for actuaries not only to understand these limitations, but to communicate them clearly to the plan sponsor. Because of the many shortcuts that the actuary takes in the process, forecast results are estimates. Each individual item (e.g., liabilities, benefit payments, assets, etc.) may differ greatly from that produced by a subsequent valuation. Because of the compensating effect of various actuarial items, however, the aggregate results obtained may still be reliable. Furthermore, comparisons of the trends of pension cost under different alternatives may be valid even when each alternative set of results is slightly off;

• **Scenario Assumptions:** The forecast results directly reflect the scenario assumptions. Because there is no certainty in the input scenario assumptions, the outcome of a forecast has a sizable margin of error. Forecast results should be presented as a range of possible outcomes. The results of a stochastic simulation, especially, should be presented in a probabilistic manner. Cost patterns should be presented probabilistically, i.e., they should communicate both the expected cost trend and the confidence level for such a cost trend through the forecast period; and

• **Forecast Report:** Because forecast results may vary by the process used, a forecast report should state clearly the methodology and assumptions, the data approximation, and other estimations employed. It is not necessary, and is often misleading, to provide detailed results for each forecast year. On the other hand, it is useful for the report to include an executive summary section that addresses the questions asked and provides concise conclusions of the study. Graphs should be used where appropriate to summarize cost trends and provide visual comparison of the alternatives.

In conclusion, forecasting is still more of an art than a science. Actuaries should not be uncomfortable about the estimations involved in the process. Even with all its limitations, however, forecasting is still one of the best tools available to help sponsors make financial decisions concerning their pension plans. Corporate executives need to make financial projections regularly, and they may find pension forecast results to be far more reliable than many of the other estimates used in corporate planning. Readers are encouraged to pursue the subject further.

In the end, forecasts are typically very exciting projects. Forecast findings usually receive much greater attention than do regular actuarial valuation results.
References


*Michael Sze*

_Hewitt Associates_

_4110 Yonge Street_

_North York, ON M2P 2B7_

_Canada_
Cost-Based Pricing of Individual Automobile Risk Transfer: Car-Mile Exposure Unit Analysis

Patrick Butler*

Abstract

Every mile traveled by a car transfers risk to its insurer. This paper posits that the product of a cents-per-mile rate based on class experience and the miles recorded on the car's odometer appropriately earns prepaid premium while the car is driven. Operation of a practical car-mile system is described briefly. To test the competing idea that driver-record pricing responds to known large differences in risk transfer, a model used to validate claim free discounts is reexamined with the car-mile as the measure of individual cost. Driver-record pricing is found to inflate car-year price-to-cost differences. Consequences of accident rate variability for a car-mile system are reviewed. The per mile cost of individual risk transfer is a class property because of the random nature of accidents. Driver-record pricing attempted on a per mile basis would amplify differences within classes.

Key words and phrases: Per mile insurance, accident rate, risk classification, driver record model, merit rating

1 Introduction

Cost-based pricing of individual risk is a key ratemaking principle promulgated by the Casualty Actuarial Society (CAS). The principle states that "A rate provides for the costs associated with an individual risk transfer;" see CAS (1993). The question for automobile insurance is how the cost of individual driving risk should be measured. When a car is not being driven, its owner has no risk to transfer for driving coverage (for all losses as a direct consequence of the car's being driven) so the cost to its insurer is zero. Every mile a car is driven adds to its risk of accident; the total cost of risk transfer increases mile by mile. Both conditions point to adoption of the car-mile (as opposed to the car-year that currently is used) as the unit of

* Patrick Butler holds a Ph.D. in geochemistry from Harvard University. The author is currently director of the National Organization for Women's insurance project and formerly was a National Aeronautics and Space Administration research scientist, principal investigator, and the Curator of Lunar Samples. He has published papers on geochemistry, lunar science, and automobile insurance pricing.
risk transfer, that is, the exposure unit. Conversion of class rates from dollars per car-year to cents per car-mile for driving coverages would be required by a one sentence amendment to rate regulatory law proposed in several states.

The purpose of this paper is to demonstrate that the car-mile exposure unit is essential to cost-based pricing of individual risk transfer. The effect of driver-record pricing on individuals is analyzed with the car-mile unit as the objective measure of cost.

2 Car-Mile Exposure Unit

The entire entry on the exposure unit in the CAS statement of ratemaking principles is: "The determination of an appropriate unit or premium basis is essential. It is desirable that the exposure unit vary with the hazard and be practical and verifiable." The currently accepted assessment of the car-mile exposure unit for automobile insurance seems to have been established by Dorweiler (1929). Regarding the variation-with-hazard requirement, Dorweiler states: "The mileage exposure medium is superior to the car-year medium in yielding an exposure that varies with the hazard, as it responds more to the actual usage of the car." Note that Dorweiler's phrase "responds more" obscures the fact that the car-year does not respond to actual use of the car. In addition, suspension of coverage during periods of no use requires administrative intervention. Dorweiler further states that "[t]he devices and records necessary for the introduction of [the car-mile] medium make it impractical under present conditions," and that while the car-year "measures the exposure prospectively, the [car-mile] require[s] a final adjustment which would be determined retrospectively."

Despite Dorweiler's assessment of superiority of the car-mile exposure unit over the car-year unit in a fundamental characteristic and his qualified judgment concerning its practicality, no substantive actuarial reassessment has been published. Bouska (1989) updates Dorweiler's paper and notes without comment that conversion to the car-mile unit has been advocated by the National Organization for Women. In a discussion of Bouska's paper, Diamantoukos (1991) observes only that the car-mile exposure unit is "perhaps a theoretically superior one in some respects" to the car-year unit.

The National Organization for Women completed a 1992 study\(^1\) for Pennsylvania legislators on operation of a car-mile system which

---

suggests that such a system would follow the odometer-limit and non-tampering conditions used in mechanical breakdown insurance policies, but otherwise would not differ much from current practice. The study follows transactions involving an example car, including a midyear sale, for four policy years. Premium payment in advance would be required to keep insurance protection in force. The premium for driving coverage at car-mile rates is prepaid in mile amounts and at times chosen by the car owner. Administrative expense and a premium for nondriving coverages are based on yearly rates and are prepaid at each policy-year renewal. Premium would be earned by the car's insurer by the day for nondriving coverages, as is currently done for all coverages, and by the mile recorded on the odometer for driving coverages. The car's insurance ID card displays the odometer-mile and date limits at which protection lapses pending further premium prepayment.

Policy renewal under this plan would be conditional on taking the car to a garage designated by the company for an annual odometer audit. The odometer would be inspected and read, and tamper-evident seals would be applied at the initial audit. Theft of insurance protection is controlled because tampering with the odometer—already a federal crime—automatically voids the policy. Driving with the cable unhooked does not steal insurance protection, because tampering usually would be detected after an accident, and tampering voids protection. The cents per car-mile rate would depend on coverage and the car's classification as appropriate by territory, use, driver, and other categories.

3 Driver-Record Pricing

Advertisements such as those promising "good rates for good drivers" lead consumers to believe that accidents can be avoided and that the important condition in individual risk of accident is how a car is driven, not how it much it is driven. This belief is encouraged through the use of merit ratings by automobile insurers to raise or lower individual prices at policy renewal time.

The actuarial literature has neglected to examine the effect of driver-record pricing on individual price-to-cost ratios where the claim rate average for the class is taken as the price and defined individual claim rates are taken as the costs of hypothetical individuals composing the class. Recent studies of driver records have

The Casualty Actuarial Society Forum (Summer 1993): 307-338. This study is available from NOW, 1000 16th Street, NW, Suite 700, Washington, DC 20036.
focused on general questions of variation in individual risk without reference to pricing or cost. For example, Mahler (1991) examines the state accident records of drivers for variation in individual risk over time (14 years), but does not discuss how the information could be applied to pricing automobile risk transfer. An earlier actuarial study done for insurance regulators, however, provides information on individual price-to-cost effects.

A widely circulated 1979 report on risk classification by insurance company actuaries on the industry Advisory Committee to the National Association of Insurance Commissioners contains a section on driver-record pricing. The report describes the issue of pricing individual risk transfer: “Many accidents are the result of chance. The problem becomes—how can insurers identify the ‘bad’ drivers from the ‘good’ drivers who were unlucky?” The impossibility of solving this problem through driver records, although downplayed in the report, is illustrated with a compound Poisson model composed of specified numbers of drivers defined to have uniform high and low annual rates of accident involvement.

In a subsequent study of driver-record pricing, Butler and Butler (1989) analyze the high and low accident rate model in terms of the car-mile exposure unit. They value the price-to-cost ratio for individual cars in terms of cents per mile and conclude that pricing based on accident, claim, or traffic violation records greatly increases the existing overpricing for unlucky owners of cars driven less than the annual average for their risk class.

Continuing justification for driver-record pricing, however, relies on the fact that cars whose drivers have had recent accidents (or traffic convictions) average more accidents in a subsequent year than do cars identically classified whose drivers have not had a recent accident. A simplified explanation for this fact—in terms of a uniform claim rate per mile—is presented below through reinterpretation of a classic model for a claim free discount plan. Assumption of a cents-per-mile cost for all cars of the model provides a base for analyzing the price-to-cost effects of driver-record pricing on individual cars. This article also considers the variation in claim rates per mile and its consequences for classification and driver-record pricing under a car-mile system in place of the assumed uniform claim rate per mile.

4 Bailey & Simon Model for Claim-Record Experience

The CAS paper “An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car,” by Bailey and Simon
(1959) is the chief reference on the CAS examination syllabus which shows and models the application of driver-record pricing to insurance for individual cars. Familiarity with its method of calculating Poisson models is required for questions on the CAS exam on advanced ratemaking; see Murdza (1992).

Bailey and Simon examine the Canadian liability claims experience of about 4 million insured car-years. The claim rate of the undivided class for each of five classes defined by car use and driver type is compared with the rates calculated for four subclasses created by sorting the records according to how many full years have elapsed since the last claim was incurred by the car's drivers.

The relative effects of sorting cars by the prior claim records of their drivers are similar for all five classes and are not affected significantly by a correction for territorial class differences. The experience for the largest Canadian class, Class 1, is shown in Table 1. The recalculated rate relative to the claim rate for the undivided class was 9 percent lower for the three year claim free subclass and progressively higher with decreasing time since the last prior claim.

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Pleasure—No Male Operator Under 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class (undivided)</td>
<td>288,019</td>
</tr>
<tr>
<td>Years Since Last Prior Claim</td>
<td>3+</td>
</tr>
<tr>
<td>Number of Claims Incurred</td>
<td>3,325,714</td>
</tr>
<tr>
<td>Car-Years Insured</td>
<td>0.087</td>
</tr>
<tr>
<td>Claims Per Car-Year</td>
<td>* Source: Bailey and Simon (1959); claim rate calculated</td>
</tr>
</tbody>
</table>

As part of their examination of the statistical justification for claim free discounts, Bailey and Simon structure a model that reproduces the decrease in the claim rate observed in the Canadian data. The model comprises cars with three annual amounts of risk transfer representing a fourfold range in annual claim rates: 100,000 cars with a uniform risk transfer rate of 0.05 claims per car-year (Amount I); 100,000 cars with a uniform rate of 0.10 claims per car-year (Amount II); and 50,000 cars with a uniform rate of 0.20 claims per car-year.
Patrick Butler Cost-Based Pricing of Auto Risk Transfer

(Amount III). The average claim rate of the model class is 0.10 claims per car-year. Bailey and Simon calculate the number of cars that would be claim free with a Poisson distribution after three years and combine them into a claim free subclass for each of the defined risk transfer rates. They calculate that the average claim rate for the new mix of the three defined rates would be 8 percent less than the class average. A subclass reduction in claim rates requires an offsetting claim-rate increase, however, to maintain the overall class average.

Because the present study concerns how all cars are affected individually by the pricing of risk transfer, the Bailey and Simon model calculations are extended here to include the subclasses with more recent prior claims. The results are compared with the Canadian experience in Figure 1. (Table 2 shows the calculated distribution of cars with the three defined risk transfer rates among the four claim-record subclasses.)

The extended model reproduces the general features of the Canadian claim data. (Bailey and Simon point out that further adjustment of model parameters would achieve more detailed agreement of the model with the Canadian data. For the present purposes, however, such adjustment would add to complexity but not to understanding.) If claim rates are taken as a measure of relative insurance prices:

- The price level for the claim free majority of cars decreases below the rate that the undivided class would pay; and
- This relatively small decrease is balanced by sharp price increases for the minority subclasses with recent claims.

The Bailey and Simon model, by reproducing empirical claim record insurance experience, shows the large variation in individual risk transfer that exists within automobile insurance price classes. Individuals in the same class are charged different prices for the same amount of risk transfer. The Amount I cars (0.05 claims per car-year) are charged four pure premiums and Amount II cars (0.10 claims per car-year) are charged two pure premiums for the same amount of risk transfer that costs the Amount III cars (0.20 claims per car-year) only one year's pure premium.

5 Risk Transfer and Miles Driven

Bailey and Simon (1960) consider reasons for the large variation in annual risk transfer within single price classes as indicated by the
Figure 1—Claim Rates of Prior-Claim Subclasses

Subclasses by Number of Claim Free Years Prior to Experience

Canadian Experience

Bailey & Simon Model
Canadian claim record experience and posited in the Bailey and Simon 1959 model for the experience. They note that driver-record and class plans are "quite ineffective in separating the better risks from the poorer risks," and conclude that:

[W]e have reached the point where we may state that the still unanalyzed cause (or causes) of variation among individual risks: (1) has a wide dispersion, (2) varies significantly from year to year for an individual risk, and (3) is measured only to a limited extent by the class plan and the merit rating plan. Annual mileage, which has long been felt to be an important measure of hazard, fits all these requirements better than any other single cause.

The first characteristic—dispersion of cars by annual miles driven—is corroborated by the U. S. Department of Transportation's nationwide personal transportation surveys. In 1977 one in five household cars was driven less than 3,000 miles, and one in ten was driven more than 20,000 miles; see Butler, Butler, and Williams (1988, p. 376).

The second characteristic—significant individual year-to-year variation in miles driven—is one that can be measured only by the car's odometer. Nevertheless, Bailey and Simon do not note a need for the car-mile exposure measure, but seem to view mileage as a lump sum class definition from which experienced car-year cost averages are used prospectively to set base price multipliers.

The third characteristic implies that variation in risk transfer amounts among individual cars resulting from differences in miles driven can be measured by class and driver-record plans. Modern class plans continue to show narrow distributions of cars by base price multiplier, in contrast to the range in miles driven; see Butler, Butler, and Williams (1988).

6 Bailey & Simon Model With Uniform Claim Rate Per Mile

Within-class variation in individual amounts of risk transfer per year can be seen as variation in the product of a rate variable and an exposure variable for each car; that is, variation in the product of a hypothetical average claim rate per mile for a car over the course of a year and the number of miles the car is driven. The current practice of charging annual rates for risk transfer implicitly assumes that the two variables cannot be resolved. In a car-mile system, however, the value of the exposure variable is recorded by each car's odometer. The following analysis of the Bailey and Simon model assumes that
all of the model cars share the same average risk-transfer rate, 0.000001 claims per mile. (The effect of presumed within-class differences in individual average claim rates per mile is considered later.) The model differences in annual risk transfer amount, therefore, are measured by the exposure variable.

The adopted claim rate per mile defines the miles per year driven for the model’s three risk amounts. For Amount I cars, 0.05 claims per year means 5,000 miles exposure per year; for Amount II cars, 0.10 claims per year means 10,000 miles exposure per year; and for Amount III cars, 0.20 claims per year means 20,000 miles exposure per year. The total risk transferred at the end of 20,000 miles traveled is the same for all cars.

<table>
<thead>
<tr>
<th>Amount of Risk Transfer</th>
<th>Miles/Year (Each Car)</th>
<th>Years Since Last Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Class (Undivided) 3+*</td>
</tr>
<tr>
<td>I</td>
<td>5,000</td>
<td>100,000</td>
</tr>
<tr>
<td>II</td>
<td>10,000</td>
<td>100,000</td>
</tr>
<tr>
<td>III</td>
<td>20,000</td>
<td>50,000</td>
</tr>
<tr>
<td>Total cars</td>
<td>250,000</td>
<td>187,594</td>
</tr>
<tr>
<td>Avg. Miles per Car-Year</td>
<td>10,000</td>
<td>9,169</td>
</tr>
<tr>
<td>Avg. Claims per Car-Year at 0.000001 Claims per Car-Mile</td>
<td>0.1000</td>
<td>0.0917</td>
</tr>
</tbody>
</table>

Table 2: Model Distribution of Mile-Amount Cars by Claim-Record Subclass

Bailey and Simon use their model to examine the mix of risks in the claim free subclass. The present study extends the analysis to obtain distributions of cars transferring the three risk amounts in the other three claim-record subclasses, as shown in Table 2. (As only the most recent claim is recognized by the plan, the claim-record distribution of the cars is calculated working back in time with a declining balance of claim free cars eligible to have a claim that counts. For example, of the 100,000 Amount I cars eligible in the 0 year, 4,877 have claims by the Poisson distribution at a 0.05 rate. The claim free balance of 95,123 cars similarly is reduced in past year 1 and so on for three years.) The miles-per-car-year average for each subclass is determined by the mix of Amount I, II, and III cars.
Although the number of cars transferring each risk amount (I, II, and III) increases with claim recency (from 2 to 1 to 0 years since the last claim), the number of highest mile cars (20,000 miles) increases most rapidly. Therefore, the average miles driven is highest (12,824 miles) in the most recent claim subclass (0 years). The average of the claim free subclass (3+ years) concurrently decreases from the class average of 10,000 miles to 9,169 miles.

7 Accidents as Random Sampling

If it is assumed that each class has uniform average claim rates per mile, automobile accidents in the Bailey and Simon model can be envisioned as a random sampling of the class population on the road. Accidents can sample only what is exposed. (Bias in the accident sampling of real car-mile class populations that results from differences in the average driving conditions encountered by individual cars is examined later in the paper.) Cars driven many miles and cars driven few miles are included in the random accident sample of the car-miles driven by the cars in the class. Because cars driven more than the class average put more miles on the road, they are overrepresented in the accident sample. Cars driven less than average are underrepresented in this sample relative to their proportion in the class. The average miles per car of the recent claim subclasses are increased through this random sampling process. The preferential selection of cars driven more miles into the recent claim subclasses also concurrently lowers (slightly) the average miles per car of the large remaining population of cars without accidents. Because of their greater average number of miles of exposure, therefore, the recent claim subclasses average more claims in a subsequent year than does the claim free subclass. All of the recent claim subclasses, however, also contain cars driven less than the class average.

8 Price-to-Cost Accuracy for Individual Risk Transfer

The miles-driven interpretation of the Bailey and Simon model provides a cost measure in car-miles for the three individual amounts of risk transferred. A price-to-cost relationship can be established for the three risk transfer amounts (I, II, and III) in the undivided class and in each of the four driver-record subclasses, a total of 15 relationships applied to the 15 groupings of cars in Table 2. (An equivalent 15 price-to-cost ratios would result from dividing the model's average claim rates per year at the five class and subclass prices by the three defined annual claim rates at the individual costs. Without being
had a recent claim, however, pay 2 cents to 3 cents more per mile
cars, which are individually driven 10,000 miles per year, is most
cent reduction in the cost per mile. Some cars in the class which have
mile amounts are different.

ums for the claim-record subclasses: $917 for the claim free subclass
(Figure 2). Claim-record subclassification transforms pricing that is
and $1,212, $1,247, and $1,282 for the progressively more recent
assumed rate of 0.00001 claims per mile produce four new pure premi­
t four subclasses created. These new mile averages multiplied by the
eds for the claim-record subclasses: $917 for the claim free subclass
percent, the adjustment would not affect the results of the analysis
miles is usually about 15 percent, the adjustment would not affect the results of the analysis significantly.)

Without claim-record pricing, all individuals pay the $1,000 per
year pure premium for the class, the same premium that Amount II
cars would pay at 10 cents a mile. At a $1,000 annual rate, however,
the 20,000 mile Amount III cars pay 5 cents a mile, while the 5,000
mile Amount I cars pay 20 cents a mile, as shown by Figure 2.

When the model class is subdivided on the basis of claim records,
the proportions of cars at the three mile amounts are changed in the
four subclasses created. These new mile averages multiplied by the
assumed rate of 0.00001 claims per mile produce four new pure premi­
for the claim-record subclasses: $917 for the claim free subclass
1212, $1,247, and $1,282 for the progressively more recent
claims subclasses. These four annual premiums divided by the three
mile amounts in each subclass produce the 12 new prices per mile for
the model cars shown in Figure 2. The effects on the cars at the three
mile amounts are different.

The effect of claim-record pricing on the risk transfer Amount II
cars, which are individually driven 10,000 miles per year, is most
telling. Without subclassification, all Amount II cars pay 10 cents a
mile for insurance. With subclassification, most of them receive a 1
cent reduction in the cost per mile. Some cars in the class which have
had a recent claim, however, pay 2 cents to 3 cents more per mile
(Figure 2). Claim-record subclassification transforms pricing that is
cost-based by definition for all Amount II individual cars to pricing that is not accurate for any cars.

It could be argued that improved price-to-cost accuracy is needed most for the model car risk transfer amounts that differ most from the class average. Without claim-record subclassification, the cars at the 5,000 mile amount pay 20 cents a mile, 10 cents a mile more than the class average price. In the claim free subclass such cars receive a 2 cent per mile reduction in price. This reduction, however, is much smaller than the 4 cents to 5 cents a mile below the class average price that the cars at Amount III (20,000 miles) pay regardless of their claim-record subclass. Furthermore, provision of this 2-cents-per-mile downward adjustment for the cars at Amount I is gained at great cost to the Amount I cars with recent claims. For these individuals, the 20-cent-a-mile amount they pay without claim-recency pricing is increased 4 cents to 6 cents a mile in the recent claims subclasses. This increase equals the entire per mile price paid by the cars at Amount III regardless of their claim-record subclass. The only negative effect for Amount III cars of pricing on claim record is that some lose a small part of their per mile subsidy (Figure 2).

Statistically, a decrease in the average cost per mile paid by Amount I cars from 20 cents to 19.3 cents coupled with an increase in the average cost per mile paid by Amount III cars from 5 cents to 5.3 cents is evidenced in a 6 percent decrease in variance of price-to-cost ratios from the three ratios of the undivided class to the twelve ratios of the driver-record subclasses. The reduced variance, however, should not mask the disparate cost of the improved statistics on individuals that is evident in Figure 2. Driver-record pricing increases the range in price-to-cost ratios paid by individuals in the same class 40 percent, from a spread of 15 cents a mile before driver-record pricing to 21 cents a mile between the lowest value for Amount III cars and the highest for Amount I cars. Operating at random on individuals, the so-called improvement increases the underpricing of risk transfer for some cars already underpriced and the overpricing of risk transfer for some cars already overpriced.
Figure 2
Effect of Model Annual Premiums on Car-Mile Prices

Class Premium $1000/Car-Year

Effective Price per Car-mile

Cost per Car-mile

Claim Record Premium ($917 to $1,282/Car-Year)

Miles Driven by Each Car
If the pricing unit were converted from car-year to car-mile so that all of the car-owners in the model class paid the same 10 cents per mile rate, however, each owner would pay only for the on-the-road protection the car consumed, while total premium received by insurers would remain the same. A car driven the model class average of 10,000 miles would experience no change in the $1,000 premium with insurance charged at car-mile rates, provided its mile amount did not change. A car driven 4,780 miles would pay $478, while a car driven 21,240 miles would pay $2,124.

9 Variation In Claim Rates Per Mile

The large differences in the type of risk environment that cars can encounter are indicated by comparing statistics for accident severities and per mile accident rates between interstate highways and city streets or between day and night driving on the same road. For example, the injury rates per million vehicle-miles of travel ranged from 0.36 on rural interstates to 3.0 on local urban roads in 1991; see Federal Highway Administration (1992). In principle, therefore, the diverse individual mixtures of car use and driving environment make it inevitable that changes in class definition would result in different claim costs per mile for new classes.

Accident rates per vehicle mile depend not only on traffic engineering classification of accidents experienced under roadway or other relevant conditions during some time period, but also on determination of the number of vehicle-miles of exposure to risk that produced the classified accidents. The same relationship holds for automobile insurance. Only if car-miles of exposure are determined can the number and cost of claims incurred within a certain time period by a certain class of cars provide any quantitative information on the expected risk transfer cost of each mile that cars in the class will travel in a subsequent rating period.

As an example of the effect of classifiable per mile differences within a business-use class of cars with adult drivers, assume two types of car use by sales representatives. With reference to the government injury rates given above, assume that one type of use covers the whole state and averages 0.25 claims per million car-miles (statewide cars), while the other covers only a metropolitan area and averages one claim per million car-miles (metro cars). Any lower average cost per claim by the metro cars resulting from lower speed urban accidents would narrow the effect on the claim cost per mile of the 4:1 claim-rate difference. Separately classifying the statewide and metro cars, provided there were enough car-miles of each usage...
type for statistical reliability, would show the differences in car-
mile cost.

10 Accidents as Biased Random Sampling

The analogy used above for viewing accidents as a process of sam-
pling car-miles on the road can be extended to presumed variations
within classes in per mile accident rates. To the extent that cars are
not classified by driver age and experience according to the known
per mile differences in accident involvement for these categories, the
accident random sampling of class car-mile populations would be
biased toward the cars driven by inexperienced drivers and by
drivers near the beginning and end of the driver age range. Further,
owing to differences in driving conditions by time and place, the
accident random sample of car-miles would be biased to the cars used
more under conditions of higher risk per mile. The accident samples,
however, also will contain cars used on average under conditions of
lower risk per mile. For example, with a Poisson distribution of
claims at the rates given for the hypothetical business use cars, 18
percent of the metro cars will incur claims in 200,000 miles of driving,
but so will 4.9 percent of the statewide cars.

11 Driver-Record Pricing on a Car-Mile Basis

Like the current driver-record pricing on a car-year basis, driver-
record pricing under a car-mile exposure unit system would have an
apparent justification in cost. The inevitable bias in an accident sam-
ple assures that the subclass of cars defined as incurring a claim in
the most-recent-miles-traveled interval—within the most recent
50,000 miles, for example—will average more accidents per mile in a
following miles-traveled interval than the class average. Applying a
recent claim surcharge to the cents-per-mile class price, however,
would constitute a deliberate, random, and unjustifiable increase in
what is paid per mile by the recent claim cars with lower than
average claim rates compared to what they would pay if they were
classified separately. Furthermore, the higher per mile charges for
the recent claim cars with significantly higher than average claim
rates per mile still would be less than what they would pay if they
were classified separately.

Because both the claim free and recent claim subclasses of a class
are mixtures of cars with above average and below average claim
rates per mile, any action to separate them must be through class
redefinition applied to the whole class.
12 Conclusion

CAS introduces its ratemaking principles with the specification that "[r]atemaking is prospective because the property and casualty insurance rate must be developed prior to the transfer of risk." In a car-mile system, evaluation of the cost per mile to be used in a prospective class rate can be done only on the basis of claim experience for a group of cars referenced to the group’s total measured car-miles of exposure that produced the claims.

What cannot be known prospectively, because it is controlled by individual car owners, is the amount of risk that will be transferred through operation of each car. Although risk transfer is paid in advance at a class rate per mile, protection is not consumed (premium is not earned by the insurer) until the risk is transferred, mile after mile, by driving. Conversely, premiums charged at car-year rates invert this cost-based relationship by charging less per mile for each mile of protection consumed, a contradiction of cost-based pricing. The assumption that this contradiction is unavoidable on practical grounds is not neutral. It favors all owners of cars driven more miles per year than the average for their class.

References


Patrick Butler
National Organization for Women
1000 16th Street, NW, Suite 700
Washington, DC 20036
Discussion of Patrick Butler's
"Cost-Based Pricing of Individual Automobile Risk
Transfer: Car-Mile Exposure Unit Analysis"

Ruy A. Cardoso*

Aside from its hyperacademic title, Patrick Butler's paper on mileage and merit rating of automobile insurance policies provides a nice twist to an old model and a reasonably compelling theoretical argument for the use of mileage as a rating variable. Yet one basic real world truth runs counter to Dr. Butler's view: automobile insurance companies generally do not use mileage as a rating variable, except in the broadest of categories. This is despite the fact that Dorweiler's justification for the use of mileage has been around for more than 60 years.

Because it generally is conceded that classification schemes have become more refined over time in response to competition, why haven't insurers already gone down the path to which Dr. Butler points? I can suppose two reasons: (1) competition doesn't really work; or (2) competition does work and the competitive market finds the use of mileage to be wanting in some respect. In my opinion, the second reason is more likely to be true.

Assuming this second reason is correct, then either the demand for or the supply of mileage rating is too low for it to be used more than it is. On the demand side, it is possible that insurance company customers don't like the notion of having their odometers inspected or of adding an uncertain level of premiums to their already complicated lives; after all, the purpose of insurance is to replace uncertain losses with certain, not uncertain, premiums. On the supply side, the costs of administering a system such as that proposed by Dr. Butler simply

* Ruy Cardoso, FCAS, MAAA, currently is vice president and chief actuary of the Automobile Insurers Bureau of Massachusetts where he represents the Massachusetts automobile insurance industry in regulatory hearings concerning insurance rates. Mr. Cardoso previously has held consulting and insurance company positions whose primary focus was the analysis of casualty insurance loss reserves. He graduated from the Massachusetts Institute of Technology in 1983 with an S.B. in management science.
may outweigh the benefits; I am unaware of any administrative cost studies that would illuminate the answer to this particular question.

Beyond pointing out this basic conflict between theory and practice, I would like to make the following observations on Dr. Butler’s analysis:

- While there is likely to be at least some correlation between variation in mileage and variation in claim frequency within a class, the Butler analysis essentially assumes a perfect correlation, disregarding the legion of unmeasurable factors that could account for as much variation as does mileage; Dr. Butler’s numerical results should be tempered considerably, therefore, before being used in the real world;

- Dr. Butler is clearly in the right when he notes that the per mile expected risk transfer cost only can be determined if real car-miles of exposure are determined. Any study based on mileage data reported by either insurers or insureds is subject to question. In the former case, this may be due to insurer indifference in reporting correct statistical data when no premium effect is involved. In the latter case, this may be due to insureds’ incentive to cheat. Here in Massachusetts, where I currently am employed, we have found that nearly 30% of policies have estimated future annual mileage of zero recorded; on the other hand, nearly 50% of policies have estimated future annual mileage of magnitudes too high to qualify for any rate discount, making it likely that these estimates are unaffected by cheating;

- Again, here in Massachusetts, we have found some evidence of a relationship between annual mileage estimates (which are based on questionable data, as explained above) and merit rating classification under the merit rating scheme used here; in particular, the higher rated (worse) drivers do tend to have higher mileage estimates, in keeping with Dr. Butler’s thesis; and

- Finally, Dr. Butler’s point (in his section 11) that “Applying a recent claim surcharge to the cents-per-mile class price, however, would constitute a deliberate, random, and unjustifiable increase” seems to argue for the complete elimination of merit rating, which the paper does not justify. As anyone who has listened to a radio talk show can attest, at least some part of the driving public demands merit rating as a way of punishing those perceived as offenders (unless, of course, the caller is one of those on the receiving end of a surcharge, in which case he or she would look on Dr. Butler’s article quite favorably). Talk show callers aside, the potential relationship between merit rating classification and other unmeasured variables (aside from mileage) cannot be dismissed based solely on this article, nor can the virtually-impossible-to-measure deference effects of a merit rating scheme.

In summary, Dr. Butler’s article, while not quite supportive of all of his conclusions, does make plain the problem of random incidence.
The principle that “cars driven more than class-average miles are over represented in the accident sample” is one that I expect many practicing actuaries frequently forget. I recall an analogous phenomenon from an undergraduate probability class; if one surveys subway riders at random and asks how many days per month they ride the subway, the average answer will be too high an estimate of the population mean because the survey-taker more likely will encounter persons who are frequently on the subway. Of course, if we all rode the subway every day, the incidence problem would go away, as would much of the need for cars and the corresponding mileage and merit rating issues. If Dr. Butler is not starting his own insurance company soon, perhaps he can devote some time to the advocacy of better public transportation systems, thereby reducing the problem he has illustrated so nicely.

Richard G. Woll*

Ruy A. Cardoso
Auto Insurers Bureau of MA
Seventh Floor
101 Arch Street
Boston, MA 02110

The purpose of this paper, according to Patrick Butler, is to “demonstrate that the car-mile exposure unit is essential to cost-based pricing of individual risk transfer.” On the basis of his demonstration, Dr. Butler advocates changing the exposure basis for private passenger automobile insurance from a car-year basis to a per mile basis. Current auto insurance prices are based on a contract that runs for a fixed period of time, usually a half year. He argues that the basis for the insurance contract for most coverages should be changed to miles driven.

Dr. Butler’s demonstration consists of creating a simplified model where there are three types of insurance customers. The first type of customer drives 5,000 miles per year. The second drives 10,000 miles per year, and the third drives 20,000 miles per year. He assumes that the risk process for each customer is Poisson with a frequency of

* Richard G. Woll is research actuary for the Allstate Research and Planning Center at Menlo Park, California. He is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries and a member of the Risk Theory Seminar of the American Risk and Insurance Association (ARIA). His work in connection with the implementation of a new auto classification plan in California included analysis of insurance results by miles driven.
one accident per 100,000 miles. For illustrative purposes, he assumes that each claim costs $10,000. He then uses this information to generate the dollars of loss experienced by each customer. This allows him to evaluate the effect of what he calls claim record pricing. This means establishing prices on the basis of prior claim records. He concludes that claim record pricing does not match prices to costs as well as charging on the basis of miles driven. He also concludes from this that "the car-mile exposure unit is essential to cost-based pricing of individual risk transfer."

Insurance companies currently recognize differences in miles driven by the use of class factors. Dr. Butler argues, however, that:

Modern class plans continue to show very narrow distributions of cars by base price multiplier in contrast to the range in the miles driven (Butler, Butler, and Williams, 1988).

Basing insurance prices on the number of miles driven makes intuitive sense. It is obvious that the difference in rates between two drivers, other things being equal, should be proportional to the difference in the miles they drive. The cost of insuring different auto customers, however, depends not only on how much they drive, but on other factors such as how well they drive, where they drive, and what kind of car they drive.

In addition, the relationship between the number of miles a customers drives and insurance claims is complex. Dr. Butler seems to assume that customers who drive more than other customers have proportionately more losses. That is, he expects a customer who drives 10,000 miles to have twice the losses of a customer who drives 5,000 miles. Allstate's data, however, present a more complicated picture. Figure 1 shows the relationship between the number of PD\(^1\) claims per mile and the number of miles driven annually by a customer. It uses information about the 1991 PD claim experience of Allstate customers in California.\(^2\)

Figure 1 shows the number of PD claims per mile going from 3.5 claims per 100,000 miles for persons who drive about 1,000 miles per year down to 0.3 claims per 100,000 miles for persons driving 30,000 miles or more. This is in sharp contrast to the constant number of

---

1 PD (property damage liability) claim frequency is used because it generally has been found to be the best indicator of underlying accident frequency.

2 Because of the passage of Proposition 103 in California which mandated the use of mileage in rating automobile policies, Allstate sent questionnaires to all its customers to get mileage data. Allstate already had collected mileage information on its customers, but the questionnaire helped to confirm the information.
claims per 100,000 miles assumed in Dr. Butler's analysis. This results in customers who drive about 1,000 miles per year having a claim frequency of 3.5 claims per year per 100 insured cars while those who drive over 30,000 miles have a claim frequency of about 8.0—a relationship of 2.25 to one, rather than the 30+ to one under Dr. Butler's assumptions.

When we turn our attention to other risk factors, we find that mileage is a relatively unimportant source of difference between customers compared to territory and years of driving experience.

The effectiveness of any auto insurance risk assessment system depends on the extent to which it matches insurance prices to insurance costs. Dr. Butler has demonstrated that the use of mileage as an exposure base in a theoretical world, where all differences in loss experience come from differences in the number of miles driven, is more effective than the use of claim record pricing. He has not demonstrated anything with respect to actual insurance experience.

The effectiveness of automobile insurance risk assessments systems was discussed extensively many years ago. A study by the Stanford Research Institute (SRI) in 1976 entitled The Role of Risk Classifications in Property and Casualty Insurance: A Study of the Risk
Assessment Process developed a means for evaluating risk assessment systems by measuring the variance of expected losses of the partitions each system produces.\(^3\)

The most efficient risk assessment system is the one that divides insurance customers into groups with the largest variance in expected losses. We also can evaluate the relative importance of various risk classification factors by measuring the percentage of the total variance each factor explains.

Dr. Butler seems to argue that the primary contributor to the variance of expected losses in the real world is the difference in the number of miles that each customer drives. There is no evidence presented by Dr. Butler, or by anyone else, to show that this is the case. The major case made for mileage in the paper is the repeated observation that insurance risk is transferred, mile after mile, by driving.

Using the SRI approach, the Allstate Research and Planning Center recently conducted a study of risk classification factors in California. The study covered most of the factors customarily used by most companies with the exception of vehicle characteristics. Allstate has collected data on the mileage driven by each customer since 1981, so the study was able to include mileage. Mileage, years licensed, and territory explained over 90 percent of the variance of the classification data included in the study for liability coverages (bodily injury liability, property damage liability, medical payments, and uninsured motorists). Over 55 percent of the total variance, however, was explained by territorial differences. Years licensed explained almost 23 percent of the variance, and mileage explained about 14 percent.

The picture was somewhat different for collision coverage. Territory, mileage, and years licensed again explained over 90 percent of the variance, but mileage explained over 33 percent of the total variance, years licensed explained about 30 percent, and territory explained about 26 percent.

---

3 The SRI report states “First, we define a measure of efficiency. Our probabilistic model for actual losses separates the random element of actual losses from the predictable element, the expected loss, that is, claim likelihood and expected claim severity. A perfectly efficient risk assessment process would be one that estimates exactly individuals’ expected losses. A process with zero efficiency would not resolve any of the initial expected loss uncertainty. A process with intermediate efficiency will be characterized by the average fraction of the initial expected loss uncertainty it resolves” (emphasis added).

The report continues that: “We find it convenient to use variance to measure uncertainty because of [its] additive property . . . In words, the expected loss variance in an entire population is equal to the sum of the average expected loss variance within each class and of the variance of the rates (average expected losses) among classes” (SRI, Supplement, p. 200).
Insurance customers with less than one year of experience have the highest losses per car. Losses per car decline each subsequent year. Thus, persons with more years of driving experience have improved loss experience. This, in turn, suggests that an important element in the transfer of insurance risk is how the customer drives. Territory rates, of course, depend on where insurance customers drive.

The Allstate study indicates clearly that how much its customers drive is only part of the overall variance of systematic risk. It is more important than the other two factors for collision insurance, but still accounts for only about one third of the total variance. It plays even a smaller role in liability insurance, the major part of auto insurance costs.

Thus, we do not believe that Dr. Butler has been able "to demonstrate that the car-mile exposure unit is essential to cost-based pricing of individual risk transfer."

Richard G. Woll
Allstate Insurance Co.
Allstate R & P Center
321 Middlefield Road
Menlo Park CA 94025-9765

Author's Reply to Discussion

The discussions by Messrs. Ruy A. Cardoso and Richard G. Woll question different points in the paper and raise other important issues concerning automobile insurance exposure units that are outside the immediate scope of the paper. Responding to these questions not only calls for expanded consideration of points discussed in the paper, but also requires examination of further consequences of conversion to the car-mile exposure unit and of retaining the car-year unit. The efforts of Mr. Cardoso and Mr. Woll in providing this opportunity and challenge are appreciated greatly.

Reply to Discussion By Ruy A. Cardoso

Mr. Cardoso's major argument against conversion to the car-mile exposure unit can be paraphrased as follows: if the car-mile were judged superior to the car-year by Dorweiler in 1929 and has not been adopted or even studied since then (over 60 years), then the car-mile unit must have some unidentified fatal flaw. Specific flaws suggested by Mr. Cardoso are (1) the technical failure of future mileage as a
classification variable, (2) the irrelevancy of exposure measurement because competition prevents overpricing, and (3) customer resistance to odometer auditing. Upon examining these suggested flaws, however, one finds evidence that the true fatal flaw that has prevented the use of the car-mile unit is seen only from the perspective of automobile insurers. Adoption of the car-mile unit as an objective standard for measuring transfer of on-the-road risk would curtail price competition severely for larger-premium consumers with broad insurance needs. It also would end the subsidy for this competition currently paid by consumers transferring less than class average risk per car-year.

Mr. Cardoso's criticism of mileage as a flawed classification variable—i.e., usable only in broadest categories, insurer indifference to integrity of data, incentive to cheat—agrees with company rate hearing testimony previously published; see Butler, Butler, and Williams (1988, p. 388). The problem with this critique is that it misses the point: the subject discussed by the current paper, as well as by the 1929 Dorweiler study, is not classification variables but exposure units. It is necessary, therefore, to clarify the difference between variables chosen to define price classes and the price unit chosen as the unit of purchase to which prices refer.

Gasoline purchase provides a ready analogy to distinguish classification variables from the price unit. Gasoline usually is available in twelve different price classes. The pricing variables that distinguish these classes are three octane levels, self service or full service, and cash or credit payment; thus, $3 \times 2 \times 2 = 12$ prices. Yet the gasoline gallon is the unit of purchase common to all of the price classes. In auto insurance, price classes are defined by variables such as territory, driver characteristics, and use of car. Distinct from such class definition variables is the price unit, currently the car-year, but which would be the car-mile after conversion to the car-mile exposure unit. Although classification variables and the price unit have distinct functions, the choices of which to use for assessing the cost of risk transfer are influenced strongly by auto insurance price competition.

In suggesting that competition currently prevents insurance overpricing of cars driven less than average, Mr. Cardoso apparently is taking the well-known fact that competition lowers auto insurance prices for marketing targets and extrapolating it to the public relations dictum that competition precludes overpricing. There is plentiful evidence, however, that insurers' price competition for customers with more risk to insure has, as its complementary effect, the overpricing of customers with less risk to insure (what Bailey (1960) calls
"skimming the cream"). This effect was described in 1911 by the New York State Legislature's Merritt Committee Report (p. 41) in its examination of the need for regulation of fire insurance pricing:

In a state of open competition the rates adjust themselves not to the hazards but largely to the strength of the insured so that the man of influence, whose patronage is desired, will get his insurance too cheaply, as against the small man who is not in a position to drive a sharp bargain. That is, competition results in discrimination.

Automobile rate hearing records contain admissions that costs are shifted from higher mileage customers to lower mileage customers and from men to women in response to price competition; see Butler, Butler, and Williams (1988, p. 405). For example, in 1982 State Farm testified to the Pennsylvania Insurance Department that in order to keep the price down for its higher mileage customers, the company keeps its low mileage discount to about half the size it should be. State Farm stated:

We're already very competitive on the [lower mileage] class, and we're generally tight on a competitive standpoint on [the higher mileage] class, and if we widen the differential, we're going to hurt ourselves very substantially on the [higher mileage] class of business.

Later in the hearing the State Farm actuary explained:

We like to follow the statistics where we can. The rating law talks about rates which are not excessive, inadequate or unfairly discriminatory, but your rating [law] also talks about doing nothing to prohibit competition in the marketplace, and as a matter of fact, we simply can't—we just can't always follow the statistical indications.

Auto insurers not only keep price differences between risk classes smaller than cost differences to compete for members of the more costly class, but also merge higher and lower risk classes or do not divide classes where such groups are distinguishable. In the latter case, for example, competition for adult men's business explains why nearly all cars in the adult driver classes are unisex-rated despite government mileage statistics, backed by accident involvement data, that show that men's average risk per year is about twice women's average risk per year. The same accident involvement data are said
to require sex-divided prices for youth classes. Rate-hearing testimony also shows that men's prices may be lowered contrary to experienced cost to allow agents to establish good relations with young men who are desirable as future sales targets.

Just as competition works to flatten rather than sharpen class differences, resistance to any real measure of exposure differences within classes also expresses competitive concern for the "man of influence" at the expense of the "small man." The capacity for miles of driving is dependent on income level, which generally determines the ability to buy gasoline and own reliable cars. Because the car-year price unit is the status quo for insurance, the result of choosing this price unit as opposed to one that responds to individual cost can be examined by analogous conversion of the price unit for gasoline from the gallon to the car-year. That is, what would the consequences be for customers if gasoline were sold like auto insurance?

With gasoline sold by the car-year, everybody with cars in the same class would pay a dollars per car-year price based on the cost per car of supplying gasoline for that class in previous years and adjusted for expected change in gasoline cost and, as currently done for auto insurance, any trend toward increased or decreased driving. Payment in advance for a car-year's worth of gasoline would allow customers to draw gasoline as needed from the class pool. Sale of gasoline by the car-year, however, would lead to problems analogous to the affordability breakdown that occurs in areas where the car-year price of auto insurance is high.

With gasoline prices set to cover the anticipated car-year average cost of each class, above average users of gasoline would experience a decrease in their gasoline expense paid by an increase in gasoline expense for below average users. Once accustomed to the benefits of unmetered gasoline, the above average user would object to any expense and accountability that using meters on gasoline pumps would entail, as Mr. Cardoso observed would occur with the use of odometers to earn insurance premiums. If the increase in annual gasoline cost per car were to force some below average users to give up cars, however, class average gallons per car-year would rise. A rise in average consumption would raise the cost of gasoline per car-year and would force still more below average users to give up their cars, causing the gasoline cost per car-year to rise even more. This death spiral effect that results when prices are not tied to a unit of individual consumption first would become apparent where the annual prices are highest, as is happening currently with auto insurance in some urban areas.
Surcharging the yearly gasoline bill of every tenth customer in a class so that the other nine can receive a customer retention discount would be analogous to the randomness of auto insurance merit rating. (Although Mr. Cardoso defends merit rating as having possible deterrence effects, customer retention is an obvious purpose. If discounts for claim free years were really risk-related, eligibility would transfer between companies. Customers generally are puzzled to discover that it does not.) With gasoline sold by the gallon instead of the car-year, however, the classification variables that set prices are certain, objective, obviously related to a cost that can be evaluated by customers, and not easily manipulated to price discriminate between customers. From the auto insurers’ viewpoint, the real fatal flaw in car-mile pricing is that it would inhibit cost shifting within classes by making the cost of individual risk transfer as understandable and controllable as the gasoline cost of automobile operation.

The public demand for driver-record pricing voiced on call-in radio talk shows to which Mr. Cardoso refers is a political response based on the only information available to consumers. Charged by the car-year, auto insurance is experienced as a flat tax on car ownership at prices based on group characteristics. By appearing to take the individual into account, driver-record pricing competes, as the paper notes, with the idea of making the car-mile the price unit for individual risk transfer.

Reply to Discussion By Richard G. Woll

Two sentences early in Mr. Woll’s discussion transform what purports to be a critique of the paper’s subject—the car-mile as the price unit for individual risk transfer—into a critique of a topic that the paper does not address—the problematic estimated future mileage discount classes with the car-year as the price unit. (These discounts are used by some insurers, but were rejected as inherently unenforceable by other insurers after several decades’ use; see Butler, Butler, and Williams (1988, p. 388)). “It is obvious,” Mr. Woll states, “that the difference in rates between any two drivers, other things being equal, should be proportional to the difference in the miles they drive. The cost of insuring different auto customers, however, depends not only on how much they drive but on other factors such as how well they drive, where they drive, and what kind of car they drive.”

While the qualifying phrase “other things being equal” in the first sentence could refer to the purpose of classifications such as those cited in the second sentence, the word “however” in the second sentence suggests a rebuttal of the first. Together they seem to imply
that the amount driven is not a measurement but a factor, i.e. a classification variable arguably related to risk, as are driver experience, garaging territory, and car type. For the remainder of the discussion, Mr. Woll criticizes the car-mile exposure unit as if it were a mileage classification variable (which it is not) to be compared with other car-year classification variables as has been done in his research at Allstate.

The basic premise of the paper is that the car-mile must work in conjunction with risk classification as the exposure unit to measure the cost of individual risk transfer. The abstract states that odometer miles multiply “a cents-per-mile rate based on class experience” and that the “per mile cost of individual risk transfer is a class property.” The essential relationship of individual exposure measurement to risk classification is emphasized in every section. It is from this perspective that the main issues raised by Mr. Woll will be addressed. These issues are within-class proportionality of cost to miles driven; observed decreasing claim rates per mile with increasing annual mileage; and car-mile costs by territory classification.

The question of proportionality of cost to miles driven is raised by Mr. Woll’s observation that Dr. Butler “expects a customer who drives 10,000 miles to have twice the losses of a customer who drives 5,000 miles.” This correctly represents how the car-mile unit works if the cars driven different distances are classified identically (and have the same coverage).

The proportionality assumed by the current car-year system, ostensibly for administrative convenience, is that within-class cost is proportional to the time period the car is insured in units of car-years. This assumption produces widely divergent per mile costs for cars identically classified. Table 1 illustrates this using Mr. Woll’s 5,000 and 10,000 miles per car-year example. The cars driven the two distances per year are garaged in the same territory and are classified identically by driver (adult unisex) and use (pleasure with limited commuting to work). The premium and per mile costs of 10,000 miles of coverage driven at 5,000 miles per car-year under two arrangements are compared with the cost of driving 10,000 miles in one car-year. Three different premiums are paid for 10,000 car-miles of exposure.
Woll's Figure 1 are obtained with data that either are unclassified or are classified only by driver sex; see Butler, Butler, and Williams (1988, p. 266). As a consequence, drivers at the extremes of the age range, who have considerably higher than average accident rates per mile and also average much less driving, would be over represented at lower mileages without classification by driver age. (The paper points out that car-miles of exposure randomly sampled by accidents would be biased toward the cars of such driver groups.) Concurrently,
the higher mileage data would be biased to cars used predominantly on limited access highways with lower accident rates per mile. As Mr. Woll points out, it is not just miles driven that determine risk transfer cost, but territory, driver, and use of car, all of which require risk classification for evaluation. Conversion of class prices from dollars-per-year to cents-per-mile demonstrates this essential relationship.

Table 2 compares the conversions of two existing car-use classes to cents-per-mile prices. All that is necessary for the conversion is an average mileage value for the class. At averages assumed for the two classes, the difference in the cents-per-mile class prices shown in the table approximate the threefold decrease in per mile claim rates with the fivefold increase in intensity of car use from 5,000 miles to 25,000 miles per year shown by the Allstate data in Mr. Woll’s Figure 1.

**TABLE 2**

<table>
<thead>
<tr>
<th>Car-Mile Prices For Two Use Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Pleasure</td>
</tr>
<tr>
<td>Business</td>
</tr>
</tbody>
</table>

* Assumed values  
** Adult unisex driver class. Multipliers from the California manual of State Farm Mutual Automobile Insurance, effective 1/15/91

What determines per-mile risk for a car is not the number of miles it is driven within an arbitrary time period (one year), but the average conditions under which the driving is done. Although intensity of car use may correlate with driver age and car use, classification is essential to determine the cost of insurance coverage per car-mile for any set of driving conditions. The car-mile unit for measuring the cost of risk transfer is also essential to meaningful territorial classification.

As though the car-mile were a classification variable, Mr. Woll states that “[W]e find that mileage is a relatively unimportant source of difference between customers compared to territory.” An example shows, however, that classification by territory depends on the car-mile exposure unit—as distinct from mileage classification—to have meaning for individual risk transfer. Table 3 shows the dollars per car-year prices for a high priced territory and a low priced territory in California for cars in the same driver and use
class. The ratio of high to low prices per car-year is 4.4, presumably representing the greater traffic density in Los Angeles and other differences in conditions and costs. The cents-per-mile costs for car owners also is shown in both territories at three mileage amounts.

### Table 3

<table>
<thead>
<tr>
<th>Car-Year Price for High Annual Mileage*</th>
<th>Car-Mile Cost to Owner by Miles Car is Driven in Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3,000**</td>
</tr>
<tr>
<td>13 Northern Counties</td>
<td>$265</td>
</tr>
<tr>
<td>Los Angeles City</td>
<td>$1172</td>
</tr>
</tbody>
</table>

* State Farm manual effective 1-15-91. Minimum coverage, adult unisex driver and car use profile from California Insurance Dept.'s 1990 Auto Premium Survey
** Discount for estimated future mileage less than 7,500 miles applied

If it is assumed that the average exposure for the class in both territories in Table 3 is 12,000 miles per car-year, conversion to the car-mile unit means that all of the northern counties cars would be paying 2.2 cents a mile and all of the Los Angeles cars in the class would be paying nearly 10 cents a mile, thus preserving the difference in territorial risk transfer costs.

In contrast to the differences between territories in cents-per-mile costs at class average mileages, the northern counties owners of cars driven 3,000 miles in a year pay more than seven cents a mile while owners of Los Angeles cars driven 20,000 miles in a year pay less than six cents a mile. The meaning of difference in risk by territory is lost if more is paid per mile for individual cars in territories with low traffic densities than is paid per mile for individual cars in territories with the highest traffic densities.

Mr. Woll devotes a considerable portion of his critique to discussing his study of statistical measures for comparing classifications of car-year data, citing evaluation methods developed by the Stanford Research Institute (SRI). Although the SRI study (1976) did not evaluate the car-mile unit as an alternative to the car-year unit, a major finding from its empirical study of nine years of individual driver accident records establishes strong limitations on the ability of classification by year to distinguish the cost of individual driving risk. The study corroborates that the most powerful class separation is driver sex, with men’s average accident likelihood per year about twice the women’s average. Despite this large class difference, how-
ever, the distributions of individual accident likelihoods per year for men and women completely overlapped, with 13% of women having likelihoods greater than men’s average and 28% of men having likelihoods less than women’s average. These overlapping distributions and averages show characteristics that are similar to the distributions of men’s and women’s annual mileages in relation to the approximately 2:1 difference in their average miles driven. Eleven percent of women exceed men’s average mileage, and 24% of men drive less than women’s average mileage; see Butler, Butler, and Williams (1988, p. 396). Individual miles of driving cannot be predicted from experienced class averages, by driver sex, or in any other way. (See the paper for the characteristics of individual mileage listed by Bailey and Simon.) The miles that individual cars are driven, however, are recorded on their odometers as the measure of individual risk transferred. The expected cents-per-mile cost of risk transfer depends on statistically reliable actual class experience.

Mr. Woll’s discussion of the car-mile price unit as if it were a classification variable has provided an opportunity to show why the car-mile exposure unit is essential to meaningful classification for individual risk transfer. Dollars-per-year prices for example risk classes that purport to distinguish differences in risk by territory, driver, and car use show large individual variability in cents-per-car-mile costs for reasons not directly related to risk. Therefore, not only is the car-mile exposure unit essential for cost-based pricing of individual risk transfer, but its use is essential in order for risk classification variables (factors) to have meaning for individual risk.

References


Merritt Committee. Joint Committee of the Senate and Assembly of the State of New York Appointed to Investigate the Affairs of Insurance Companies Other Than Those Doing Life Insurance Business, Report, Assembly No. 30 (1911).

A New Approach to Modeling Excess Mortality

Peter D. England and Steven Haberman

Abstract

This paper describes a complete framework for the statistical modeling of excess mortality, with particular reference to the experience of insured, impaired lives. The principal measure of excess mortality considered is the standardized mortality ratio. The modeling approach, based on the theory of generalized linear models, allows us to build models containing several explanatory variables. The statistical significance of these variables can be tested, and the effect of interactions between the variables can be assessed rigorously. The paper uses data drawn from the extensive, continuing investigation into the mortality of insured, impaired lives conducted by the Prudential Assurance Company (UK). The methodology has close connections with the traditional actuarial approach to the measurement of excess mortality and can be regarded as a generalization of this traditional approach.

Key words and phrases: impaired lives, generalized linear models, multiplicative hazard

1 Introduction

1.1 Background

In 1947, the Prudential Assurance Company decided to institute an inquiry into the mortality experience of medically impaired, insured lives. The investigation was designed to be both medical and actuarial. The data were drawn from holders of life insurance policies effected since July 1947 in the ordinary branch of the Prudential Assurance Company. Policies were included if the life insured exhibited one of a long list of impairments. Lives exhibiting two or more

---

1 Peter D. England obtained his Bachelors degree in Actuarial Science in 1988 and then stayed at City University, London, to assist in the research activities of the Department of Actuarial Science and Statistics. In 1993 he completed his Ph.D. in statistical modeling of excess mortality. He current is working for Commercial Union plc and specializes in non-life insurance.

2 Steven Haberman is Professor of Actuarial Science and head of the Department of Actuarial Science and Statistics at City University. He received a degree in mathematics from Cambridge University. He joined the Prudential Assurance Company as an actuarial trainee and then City University as a lecturer, qualifying as an FIA in 1975. He has published widely in actuarial and related fields. His current research interests include mortality, morbidity, premium rating, and pension funding.
major impairments were excluded from the investigation, where an impairment was regarded as major if it would warrant a surcharge in its own right.

It was not considered practical to extend the scope of the investigation to include every impairment encountered in the course of underwriting. For impairments that occur comparatively infrequently, sufficient data would not have accumulated to provide useful results. Accordingly, the rarer conditions generally have been excluded. At the outset it was not possible to foretell the quantity of data that would be forthcoming, however, and certain groups were included where experience has shown that the data have proved insufficient.

Since 1961, several authors have reported results based on the Prudential impaired lives data set. A comparison of the diverse reports is informative and provides insight into the changes in excess mortality over the 40 years that the investigation has been operative. It is worth considering the difference in the scope of the studies and the approach adopted by the various authors (Table 1).

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication Date</th>
<th>Calendar Years of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarke</td>
<td>1961</td>
<td>1947 to 1958 (All Impairments)</td>
</tr>
<tr>
<td>Preston &amp; Clarke</td>
<td>1966</td>
<td>1947 to 1963 (All Impairments)</td>
</tr>
<tr>
<td>Clarke</td>
<td>1979</td>
<td>1964 to 1973 (All Impairments)</td>
</tr>
<tr>
<td>Leighton</td>
<td>1987</td>
<td>1974 to 1983 (All Impairments)</td>
</tr>
<tr>
<td>Papaconstantinou</td>
<td>1988</td>
<td>1947 to 1981 (All Impairments)</td>
</tr>
<tr>
<td>Renshaw</td>
<td>1988</td>
<td>1947 to 1981 (Hypertension)</td>
</tr>
<tr>
<td>Haberman and Renshaw</td>
<td>1990</td>
<td>1947 to 1981 (Peptic Ulcer)</td>
</tr>
</tbody>
</table>

The studies by Clarke (1961), Preston and Clarke (1966), Clarke (1979), and Leighton (1987) form a series in which the authors use the same approach in their analyses. Traditional methods were used to produce standard actual/expected (A/E) ratios only. The differences between the reports lies in the exposure-to-risk periods considered (Table 1) and in the control experiences used in the calculation of expected deaths. These authors briefly comment on the excess mortality of female lives where there are sufficient data to provide useful results.

---

3 The Prudential data set is not freely available. Readers interested in obtaining this data set should direct inquiries to Prudential Assurance, Holborn Bars, London, ECIN 2NH, England.
Papaconstantinou (1988) uses the entire data set (as available at the time) in his analysis and uses conventional exposure-to-risk theory in the calculation of mortality rates. He uses the data available to provide a comprehensive set of excess mortality measures including not only the familiar A/E ratios, however, but also excess death rates and measures based on cumulative mortality. He considers all impairments (male and female combined) for which there are more than 100 entrants.

The Prudential impaired lives data set first was used in the statistical modeling of excess mortality by Renshaw (1988) who adopts the multiplicative hazards approach. Renshaw (1988) and Haberman and Renshaw (1990) use the same data as Papaconstantinou (i.e., data for 1947 to 1981) and provide results for two impairments, hypertension and peptic ulcer (male lives only), respectively.

1.2 Summary

This paper concerns the measurement of excess mortality experienced by impaired, insured lives. The approach adopted here is to use a multiplicative hazards model for the force of mortality. This is similar to that used by Renshaw, but additionally includes data for the period 1982 to 1987. The methodology is described and illustrated with examples drawn from the Prudential impaired lives data set. It must be mentioned that this approach has been applied systematically to all of the major impairment groups in the Prudential study, and the full results are given in England (1993).

The methodology used in this paper can be applied to any investigation of excess mortality if the data requirements can be satisfied. Such investigations would include studies based on the experience of a single company or comprising the pooled experience across a number of companies. An example of such an investigation is the Medical Impairment Study 1983 of the Society of Actuaries and Association of Life Insurance Medical Directors. We feel that, given the power and versatility of the methodology, its potential should be recognized outside the United Kingdom (where it has been applied so far). Given their ready access to fast personal computers, workstations, and mainframe computers, we feel that this method will be of interest to North American actuaries.

The results of this study support and supplement the results published in earlier investigations. The results relating to the subsidiary impairment codes are new—this information has been ignored in earlier investigations based on the Prudential data set. These latest
England and Haberman

analyses identify some anomalous results emerging from the earlier studies, in particular some of the indices given by Papaconstantinou (1988).

1.3 A Note of Caution

When comparing results of the various studies, differences in the mortality ratios obtained may be due to combinations of the following four factors:

- **The Period Under Study**;
- **The Control Experience Used in the Calculation of Expected Deaths**: It should be noted that when expected deaths are low, a small difference in the value of expected deaths may change the value of the mortality ratio significantly, as expected deaths appear in the denominator;
- **The Method Used in the Calculation of Expected Deaths**: Using traditional methods, expected deaths are given by an expression of the form $Eeq^*$ (or $\Sigma Eeq^*$) i.e., the exposure to risk multiplied by the standard mortality rate. Using the multiplicative hazards approach, expected deaths ($e_j$) are given by a different expression based on the aggregate integrated standard force of mortality. Differences may arise in the values of expected deaths given by these methods. When expected deaths are low, these differences may cause a significant change in the value of a mortality ratio, as expected deaths appear in the denominator; and
- **Errors**: Despite the efforts taken to eliminate any source of error, it is possible that errors occur that affect the results obtained, especially in a study the scale of an impaired lives investigation. Errors may be due to incorrect recording of data, mistakes in data manipulation, programming mistakes, incorrect calculations using results, and typographical errors in reports. Major errors usually are immediately noticeable; minor errors, however, may pass undetected.

2 The Data Set

The 1947 Prudential impaired lives study uses a coding scheme for impairments devised by the company’s principal medical officer at the time, T.W. Preston. The impairments considered are divided into nine broad categories (e.g., circulatory impairments, respiratory disorders), each subdivided into its constituent impairments. Since 1947, some impairments that originally were included have been
dropped and some impairments that were not have been added. A few impairments have had their associated codings changed. By the end of 1987, data were available on over 650,000 policies effected on impaired lives (where the impairment was present at the outset).

Those involved in planning the study showed considerable foresight and adopted a classification of impairments that was criticized in its day for being too detailed. To this criticism the powerful riposte was made that "once the data have been tabulated, groups can always be combined but they can never be further subdivided." It is only now, with sufficient data and statistical software packages, that full advantage of the detailed classification can be made.

For each policy in the investigation, the following information was recorded: policy number, impairment code (plus subsidiary code), date of entry (year and month), age at entry (next birthday), date of exit (year and month), age at exit (next birthday), curtate duration at exit, mode of exit (still in force, withdrawal, death), cause of death, joint life marker, and sex.

Information that would be of interest, but which is not available, concerns the terms of acceptance (accepted as standard, reducing debt etc.), duration since onset of impairment, sum insured, type of policy, experience of lives declined for insurance, and smoking status.

3 Statistical Methodology

3.1 Traditional Methods

The traditional actuarial approach to the measurement of mortality is based on the comparison of actual and expected deaths. The history of this process has been investigated by Keiding (1987). This approach also has been applied to the measurement of excess mortality associated with an extra risk in the comparison of actual and expected deaths for a group of policyholders exhibiting the particular risk under consideration. Examples of possible types of risk are medical impairments, occupational hazards, hazardous pursuits, geographical location of residence, and ethnic origin. Combinations of the above risks may be of interest (e.g., the effect of a particular disease within different ethnic communities).

Using exposure-to-risk theory, the expected numbers of deaths are calculated using a set of suitable standard mortality rates, controlling

---

4 One of the earliest descriptions of the method is attributable to William Dale, an English actuary living in the 18th century who was investigating the adequacy of contemporary annuity rates.
England and Haberman Excess Mortality

as closely as possible for factors such as sex and age (and possibly other characteristics). Let \( d_t, q_t, E_t \) be the observed number of deaths, the observed mortality rate, and the initial exposure-to-risk, respectively, for the group under consideration for the interval of follow-up for curtate duration \( t \) (i.e., \([t, t+1]\)) where \( t \) is an integer (say, measured in years). Note that \( d_t \) and \( q_t \) are random variables. Let \( q_t^* \) be the standard mortality rate and define \( d_t^* \) to be the expected number of deaths, i.e., \( d_t^* = E_t q_t^* \).

The interval mortality ratio for the interval of follow-up \([t, t+1]\) is denoted by \( k_t \), and is given by:

\[
k_t = \frac{d_t}{d_t^*} = \frac{q_t}{q_t^*}.
\]

Clearly, if \( k_t > 1 \), the mortality rate in the study group (for curtate duration \( t \)) is higher than the standard rate. If \( k_t < 1 \), the mortality rate in the study group is lower than the standard rate.

When the numbers of deaths (or expected deaths) are low, neighboring intervals of follow-up sometimes may be grouped together to give:

\[
nk_o = \frac{\sum_{t=0}^{n-1} d_t}{\sum_{t=0}^{n-1} d_t^*}
\]

where neighboring intervals are grouped over an \( n \) year period. Thus, \( nk_o \) is the ratio of deaths observed in an \( n \) year period and deaths expected over the same period.

The properties of this ratio have been described in detail by Haberman (1988). For example, it is straightforward to show that the index \( nk_o \) may be regarded as a weighted average of the \( k_t \) over the first \( n \) years of observation with weights \( w_t \) equal to the number of deaths expected on the basis of the standard population mortality rates at duration \( t \). The index clearly does not treat all the \( k_t \) terms equally; it places most weight or emphasis on those \( k_t \) at the durations where the \( w_t \) are highest. This may not be unreasonable; these durations are likely to be those where the underlying data are largest. Any resulting indices are subject to the least statistical variability and, hence, are most reliable in statistical terms. The quantities \( k_t \) and \( nk_o \) commonly are known as the A/E ratios; see Clarke (1979).
This conventional approach (based on A/E ratios) does not provide any guidance on how we decide which factors or combinations thereof have a significant bearing on excess mortality, nor how we should construct the best possible model representing excess mortality. For such refinements, we need a more statistically sound structure.

3.2 The Multiplicative Hazards/Generalized Linear Model Approach

3.2.1 Introducing the Multiplicative Hazards Model

Following the notation used in the literature on survival analysis (including Elandt-Johnson and Johnson, 1980; London, 1988), we let the random variable $T$ denote the lifetime of a living organism or an inanimate object (e.g., a light bulb). The instantaneous failure rate at time point $t$ is $A(t)$ and is called the hazard rate or intensity rate. Strictly,

$$A(t) = \lim_{\Delta t \to 0^+} \frac{\text{Prob}(t < T < t+\Delta t \mid T > t)}{\Delta t}$$

where this limit is assumed to exist. Hence, the probability of failure in $(t, t + \Delta t)$ given survival to time $t$ is approximately equal to $A(t) \Delta t$, for very small $\Delta t$.

Now consider the hazard rate of a study group with certain characteristics $(z)$ and denote this hazard rate by $A(t,z)$. Note that $z$ is a vector of information on the characteristics of the study group. If the important characteristics are age, sex, weight, height, and impairment, for example, then the vector $z$ may be as follows: $z = (50$ years, male, $250$ pounds, $5'10''$, hypertension).

The multiplicative hazards model is said to hold when $A(t,z)$ can be factored as

$$A(t,z) = A^*(t) \times \theta_z$$

where $A^*(t)$ is some known standard hazard rate, independent of $z$, and the proportionality factor $\theta_z$ (independent of $t$) measures the effect of the characteristic $z$ on the study group's hazard rate relative to the known standard hazard rate $A^*(t)$. If $\theta_z > 1$, the failure rate in the study group is greater than the standard failure rate, and if $\theta_z < 1$, the failure rate in the study group is less than the standard failure rate.
In the field of actuarial science, failure is (typically) death, time of failure is age at death, and the hazard rate is called the force of mortality. Thus, equation (1) can be rewritten by replacing the symbol for the hazard rate ($\lambda$) by the standard symbol for the force of mortality ($\mu$) and rearranged to give:

$$\theta_z = \frac{\mu(t,z)}{\mu^*(t)}.$$  

In this context the $\theta_z$'s can be viewed as instantaneous mortality ratios. These ratios can be compared with the interval mortality ratio, $k_t$, which is a ratio of annual mortality rates.

Cox (1972) proposes writing the proportionality parameter $\theta$ as an exponential function of the vector of covariates $z$ with unknown regression parameters $\beta$ such that:

$$\theta_z = e^{\beta^T z}$$ \hspace{1cm} (2)

giving

$$\mu(t,z) = \mu^*(t) e^{\beta^T z}.$$ \hspace{1cm} (3)

This representation of the mortality ratio $\theta_z$ is a mathematical construct. It uses only those factors that are considered to influence the mortality ratio to a significant extent. These factors may be qualitative, such as severity of disease, or quantitative, such as age at entry or level of blood pressure. Both $\beta$ and $z$, however, must be real-valued vectors. To accomplish this, nominal and ordinal characteristics are usually coded using real numbers.

Equation (1) is called the multiplicative hazards model or the log-linear model because the linear combination of factors acts multiplicatively on the mortality ratio. This equation provides a specific case of a more general model:

$$\theta_z = h(\beta^T z)$$

where $h$ is a function to be specified. Detailed experiments with different choices for $h$ have shown that the exponential function provides the most satisfactory choice in terms of the goodness of fit, its simplicity, and its implicit avoidance of negative values for $\theta$. 

92
3.2.2 Estimating the Parameters $\beta$

Consider a study consisting of $N$ mutually independent individuals. For the $i^{th}$ individual, $i = 1, 2, \ldots, N$, let the entry and exit times from the study be denoted by $\tau_i$ and $t_i$, respectively. Let $z_i$ be the vector of characteristics associated with $i$, and let the indicator variable, $\xi_i$, denote the mode of exit, i.e., $\xi_i = 1$ if the $i^{th}$ person died in the study and $\xi_i = 0$ otherwise. This results in the following likelihood function:

$$L = \prod_{i=1}^{N} \mu(t_i,z_i)^{\xi_i} \frac{S(t_i,z_i)}{S(\tau_i,z_i)}$$

where:

$$S(t,z) = \exp[-\int_0^t \mu(s,z) \, ds]$$

is the survival function. The log likelihood function may be written as:

$$\log L = \sum_{i=1}^{N} \left( \xi_i \log \mu(t_i,z_i) + \log S(t_i,z_i) - \log S(\tau_i,z_i) \right)$$

$$= \sum_{i=1}^{N} \left( \xi_i \log \mu(t_i,z_i) - \int_{\tau_i}^{t_i} \mu(u,z_i) \, du \right). \quad (4)$$

Substituting Cox’s multiplicative hazard function (represented by equation (3)) into equation (4) gives the following result:

$$\log L(\beta) = \text{constant} + \sum_{i=1}^{N} \left( \xi_i \beta^T z_i - e^{\beta^T z_i} \int_{\tau_i}^{t_i} \mu^*(u) \, du \right) \quad (5)$$

where we identify specifically the dependence of $\log L$ on the unknown regression parameters, $\beta$.

For convenience, the individuals in the study are grouped into $M$ cohorts, denoted by $j \ (j = 1, 2, \ldots, M)$, which represent particular combinations of the characteristics under consideration—for example, age at entry, policy duration, and severity of the impairment. Renshaw (1988) shows that, in this case, the log likelihood function may be written as:
\[ \log L(\beta) = c + \sum_{j=1}^{M} [d_j \log(m_j) - m_j] \]  \hspace{1cm} (6)

where:

- \( c \) = a constant independent of \( \beta \);
- \( N_j \) = number of individuals in cohort \( j \) so that \( N = \sum_{j=1}^{M} N_j \);
- \( t_{jk} \) = age at which individual \( k \) from cohort \( j \) was last observed;
- \( \tau_{jk} \) = age at which individual \( k \) from cohort \( j \) entered the study;
- \( d_j \) = number of deaths observed in cohort \( j \);
- \( e_j \) = the aggregate integrated force of mortality, 

\[ e_j = \frac{N_j}{\sum_{k=1}^{N_j} \int_{\tau_{jk}}^{t_{jk}} \mu^*(u) \, du}. \]  \hspace{1cm} (7)

In the appendix, it is proved that \( e_j \) is an unbiased estimate of the expected number of deaths from cohort \( j \) given standard rates had applied. The interpretation of \( e_j \) is discussed in the appendix; and

\( m_j \) is given by:

\[ m_j = e_j \times e^{\beta^T z_j} \]  \hspace{1cm} (8)

so that

\[ \log(m_j) = \log(e_j) + \beta^T z_j. \]  \hspace{1cm} (9)

Note that equation (6) has the same form as the log likelihood for independent Poisson random variables \( d_j \) with respective means \( m_j \), given by equation (8).

The vector of parameters \( \beta \) can be estimated by maximizing the log likelihood function. This has been performed using the software package GLIM, which relies on the presence of a generalized linear model. The above description can be recast in the form of a generalized linear model, as discussed by Dobson (1989).

Equation (9) is used as the estimation equation in the GLIM package. The term \( \beta^T z_j \) is called the linear predictor in generalized linear model terminology. The log \( (e_j) \) term is called an offset and is considered as an extra term in the linear predictor with a coefficient of 1. GLIM calculates the parameter estimates using maximum likelihood techniques.
3.3 Practical Implementation

To establish a connection between the factors \( e^{\beta T z_j} \) and traditional actuarial mortality ratios, recall equation (8) (rewritten in the form):

\[
e^{\beta T z_j} = \frac{m_j}{e_j}.
\]

Following the methods used in the appendix, it easily can be proved that \( m_j \) is an unbiased estimate of \( E[d_j] \). It thus seems reasonable to replace \( \beta, m_j, \) and \( e_j \) by \( \hat{\beta}, d_j, \) and \( d_j^* \), respectively, where \( d_j^* \) is the expected number of deaths in cohort \( j \) had standard mortality rates applied. This gives

\[
e^{\hat{\beta} T z_j} \approx \frac{d_j}{d_j^*} = \frac{Actual \ Deaths}{Expected \ Deaths}.
\]

which is identical to the traditional actuarial A/E mortality ratio.

In practice, the application of this statistically-based methodology is straightforward. For a mortality study involving \( N \) participants, the first step is to partition the sample into homogeneous cohorts (indexed by the suffix \( j \)), ensuring that there are sufficient data in each cohort to make the construction of mortality ratios meaningful. For each individual, the information needed for the statistical calculations is age at entry (\( r \)), age at exit (\( t \)), and mode of exit. The second step is to use this information to calculate the observed number of deaths (\( d_j \)) and the expected number of deaths (\( e_j \)) for each cohort. The third step is to develop the coding scheme used to identify the covariate structure to be modeled. It should be emphasized that the covariates must be expressed as real-valued variables. The fourth step is to enter the values of \( d_j \) and \( e_j \) and the covariate structure into a statistical modeling software package, such as GLIM, for model fitting and the calculation of the parameter estimates, \( \hat{\beta} \).

Using GLIM, various models may be fitted from the null model (no covariates) to more complex models involving interaction terms, giving different parametric representations for the mortality factor. Statistical analysis of the significance of covariates and their possible interactions is based on a goodness-of-fit statistic called the deviance. (GLIM automatically provides the deviance when fitting models). The deviance is based on the likelihood ratio principle rather than on the (possibly more familiar) Pearson goodness-of-fit statistic. It is essential that inferences should be based on differences
between model deviances, as their absolute values are conditional on the total number of covariates under simultaneous investigation. The differences in model deviances are assumed to follow the $\chi^2$ distribution with the appropriate degrees of freedom (an approximate result). This can be used to assess the significance of factors included in (or excluded from) the model. Furthermore, residual plots may be used as an informal diagnostic tool to highlight the source of unexpected effects.

If a model provides a good fit, the histogram of deviance residuals should be approximately bell-shaped (i.e., approximately normal). Also, a scatter plot of deviance residuals against linear predictor should show a corridor of values.

Any other patterns would be indicative of a lack of fit. In such a case, a transformation of the data may be necessary, or account may need to be taken of factors other than those included in the current model. Outliers also would be detected by plotting residuals and would be identified as isolated points on these plots far from the remaining residuals.

This methodology is similar to that used in cohort studies in epidemiology, which are concerned with the follow-up of large population groups over many years (for example, to ascertain the effects of environmental exposures on the outbreak of illness). A full description is given by Breslow (1985). One of the main differences between the two approaches is that in this paper we are modeling the relative mortality experience, whereas in epidemiological studies the mortality rate itself often is modeled.

### 3.4 Basis for Expected Deaths—The Standard Experience Used

Choosing a suitable control experience for the calculation of expected deaths is a difficult task. The ensuing discussions of papers presented to the Institute of Actuaries concerning the mortality of impaired lives indicate that criticism often rests with the choice of the control experience.

One of the principal problems is with the length of the investigation, presently 41 years. In his analysis, Papaconstantinou (1988) modifies the A67-70 (2) select table using linear relationships proposed by the Continuous Mortality Investigation Bureau\(^5\) to produce a

\(^5\) The A67-70 (2) select table is a standard life table with a two year select period based on data (male endowment and whole life policyholders) collected by the Institute and Faculty of Actuaries Continuous Mortality Investigation Bureau (CMIB) from contributing insurance companies (Joint Mortality Investigation Committee, 1974). The CMIB is a permanent research organization established by the Institute and
different set of mortality rates for each quadrennium between 1949 and 1978. By a process of interpolation and extrapolation, he produces a different set of rates for male lives for each year of entry 1947 to 1981. For female lives, he uses the rates thus produced with a four year age deduction. Renshaw (1988), in his turn, feels that the method adopted by Papaconstantinou is unnecessarily detailed and condenses Papaconstantinou’s rates into five year intervals commencing with 1947 to 1951 and ending in 1977 to 1981.

In the analysis covered by this report, it was decided to use the A67-70 (2) select table unmodified for all years of entry for male lives. The period used in forming this table (1967 to 1970) is roughly mid-way through the period for which the Prudential data are available (1947 to 1987). The expected deaths calculated for the earlier part of the study will tend to be understated (resulting in an overstatement of the excess mortality). Similarly, the expected deaths calculated for the later part of the study will tend to be overstated (resulting in an underatement of the excess mortality). A comparison of Renshaw’s results, in respect to hypertension, with the results included in this report, however, reveals that the differences in the standard forces of mortality used, on the whole, make little difference to derived measures of excess mortality.

Nevertheless, the basis for expected deaths used here is not ideal. A more satisfactory approach would be to obtain the Continuous Mortality Investigation Bureau’s data for whole life and endowment insurances (for standard lives), grouping into suitable time intervals (e.g., four or five years), and graduating to form a smooth set of mortality rates for each time interval.

The use of a fixed control experience does not mean that time trends are being ignored. Any significant trends would be identified through the presence of calendar year of death as one of the covariates in the vector $z$.

4 Illustrating the Methodology

To illustrate the methodology and highlight some of the advantages of the multiplicative hazards/generalized linear model approach, a summary of the results from preliminary analyses of two impairment groups (impairment of coronary arteries and hypertension) will be considered (male lives only).

First, consider the data set. Of the data available for each policyholder, the information needed is medical impairment (including

Faculty of Actuaries to collect and analyze mortality and morbidity statistics and prepare standard tables. The continuous collection of such data began in 1924.
further subclassification), date of entry, age next birthday at entry, date of exit, mode of exit (withdrawal, death), and sex. Medical impairment and sex provide the necessary information for breaking the sample into reasonably homogeneous cohorts. The age on next birthday at entry provides the necessary values $t_{jk}$. Date of entry, age next birthday at entry, and date of exit together provide the values of $t_{jk}$ needed for the calculation of $e_j$ (using equation (7)). The values of $d_j$ depend on the mode of exit.

4.1 Impairment of Coronary Arteries

The initial selection and sorting of data were carried out using the SPSSX statistical software package. For both male lives and female lives separately, a subset of the full data set was created that includes only those lives identified as suffering from impairment of coronary arteries. This category includes thrombosis, occlusion, ischaemia, infarction, and angina.

The data are partitioned according to:

- **Age at Entry**: taking four levels
  - (1) 16 to 39
  - (2) 40 to 49
  - (3) 50 to 59
  - (4) 60 to 79;

- **Policy Duration**: taking three levels
  - (1) 0 to 2 years
  - (2) 2 to 5 years
  - (3) 5 to 8 years; and

- **Whether Complications are Present** (complications defined as subsequent chest pain on exertion): taking two levels
  - (1) Without complications
  - (2) With complications.

This gives a total of 24 cohorts ($4 \times 2 \times 3$). For each of these cohorts, the number of deaths observed ($d_j$) and expected deaths ($e_j$) are calculated (using Fortran 77 programs specially written for this purpose).

Age at entry and policy duration are taken as discrete variables. It would be possible to use age and duration in continuous form and represent their effect on excess mortality through the use of an appropriate (regression) model. In the results reported here, this approach is not adopted.

Once the data are partitioned according to the covariate classification chosen and the number of deaths observed and expected (the aggregate integrated standard force of mortality) are calculated for each cross-classified cohort, the data are suitable for feeding into the GLIM software package for model fitting and statistical analysis. Here equation (9) is used as the estimating equation. The method of model fitting adopted is forward stepwise, i.e., start with the simplest model (the null model) and include parameters one by one.
4.1.1 The Null Model

The null model has the simplest form of structure in which the linear predictor is represented by a single parameter, i.e.,

$$\beta^Tz = b.$$ 

This is equivalent to combining age at entry, duration, and complication groups to give an overall mortality ratio equivalent to

$$\frac{\text{Total Deaths in Study Group}}{\text{Total Expected Deaths}}$$

which is an estimate of the mortality ratio associated with impairments of the coronary arteries as a whole.

Fitting this model with GUM gives a parameter estimate

$$\hat{b} = 0.9076.$$ 

From equation (11), the mortality ratio is given by $e^{\hat{b}^Tz}$, giving

$$e^{0.9076} = 2.48.$$ 

Thus, the overall mortality ratio for life insurance policyholders with impairments of the coronary arteries at entry is 248% (i.e., extra mortality of +148%).

4.1.2 Main Effects Models

More information can be obtained by fitting models that allow for inclusion of the principal factors believed to influence excess mortality (i.e., age at entry, duration, complications). These factors are called main effects to distinguish them from the interaction terms that relate to interdependence between factors. In this section, we shall consider models that include these main effects, fitted separately.

4.1.2.1 Age at Entry

Recall that only the age at entry data are partitioned into four levels. We denote the parameter values associated with the level of age at entry as $\alpha_i$. Therefore, the parametric representation of the linear predictor for the $i^{th}$ cohort is:
\[ \hat{\beta} T_{z_i} = b + \alpha_i \quad i = 1, 2, 3, 4. \]

Fitting the age at entry main effects model in GLIM gives the following parameter estimates:

\[ \hat{b} = 2.912, \quad \hat{\alpha}_1 = 0, \quad \hat{\alpha}_2 = 1.236, \quad \hat{\alpha}_3 = -1.679, \quad \hat{\alpha}_4 = -2.471 \]

(For technical reasons, the first parameter estimate for any factor included in a model is assigned the value zero).

Calculating \( e^{\hat{\beta} T_{z_i}} \) for each \( i \) gives:

<table>
<thead>
<tr>
<th>Age at Entry</th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 to 39</td>
<td>1839</td>
</tr>
<tr>
<td>40 to 49</td>
<td>534</td>
</tr>
<tr>
<td>50 to 59</td>
<td>343</td>
</tr>
<tr>
<td>60 to 79</td>
<td>155</td>
</tr>
</tbody>
</table>

These results indicate that proposers for life insurance under 40 year of age suffering from impairments of the coronary arteries constitute a substantial extra risk. Excess mortality decreases as age at entry increases. Lew and Gajewski (1990), in their review of excess mortality experience, note that for most medical impairments, relative mortality is highest at ages under 40 and decreases with advances in age to relatively low mortality indices at ages 60 and over.

4.1.2.2 Policy Duration

Recall that the policy duration data are partitioned into three levels. We denote the parameter values associated with the \( j^{th} \) level of policy duration as \( \delta_j \). Therefore, the parametric representation of the linear predictor for the \( j^{th} \) cohort is:

\[ \hat{\beta} T_{z_j} = b + \delta_j \quad j = 1, 2, 3 \]

Fitting the policy duration main effects model in GLIM gives the following parameter estimates:

\[ \hat{b} = 1.205, \quad \hat{\delta}_1 = 0, \quad \hat{\delta}_2 = -0.4727, \quad \hat{\delta}_3 = -0.4073 \]

resulting in the following mortality ratios:
These results show that the mortality ratio in the first two years after entry is higher than subsequently. Such results also are not unexpected. Lew and Gajewski (1990) also comment that indices of excess mortality tend to decrease with increasing policy duration.

4.1.2.3 Complications

Similarly, we denote the parameter values associated with the $k^{th}$ level of the presence of complications as $\gamma_k$. Therefore, the parametric representation of the linear predictor is:

$$\beta^T z_k = b + \gamma_k \quad k = 1, 2.$$

Fitting the complications main effects model in GLIM gives the following parameter estimates:

$$\hat{b} = 0.7893, \quad \hat{\gamma}_1 = 0, \quad \hat{\gamma}_2 = 0.2771$$

resulting in the following mortality ratios:

<table>
<thead>
<tr>
<th></th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Complications</td>
<td>220</td>
</tr>
<tr>
<td>With Complications</td>
<td>290</td>
</tr>
</tbody>
</table>

As expected, there is a higher risk associated with the presence of complications.

4.1.3 Significance of Main Effects

The results according to the main effects fitted separately could have been obtained using the traditional actuarial methods, based on A/E ratios. One of the advantages of the modeling approach used in this paper, however, is that it is now possible to assess the statistical significance of the main effects. That is, it is possible to answer such questions as “Is age at entry a significant rating factor?” and “What about the presence or absence of complications?” These questions are answered with recourse to the model deviances. The null model is a simpler model than the main effects models, and it can be
shown that the difference in deviances between the null model and the main effect models approximately follows a $\chi^2$ distribution; see Dobson (1989).

Using the deviances provided by GLIM when fitting the particular models, a deviance table for the main effects models may be produced as in Table 2 below. The differences in model deviances are referred to the appropriate $\chi^2$ distribution to assess the significance of the main effects.

**Notation**

The null model is denoted by $H_0$, the age at entry model by $A$, the policy duration model by $D$, and the complications model by $C$.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Impairment of Coronary Arteries: Significance of Main Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Deviance</td>
</tr>
<tr>
<td>$H_0$</td>
<td>116.56</td>
</tr>
<tr>
<td>$A$</td>
<td>33.75</td>
</tr>
<tr>
<td>$D$</td>
<td>102.27</td>
</tr>
<tr>
<td>$C$</td>
<td>110.95</td>
</tr>
</tbody>
</table>

Analysis of the differences in model deviances indicates that all three main effects are highly statistically significant (tail area less than 5% in all cases).

**4.1.4 More Complex Models**

Because all three main effects are significant, we may be interested in more complex models, looking at age at entry and policy duration combined or including all three factors. We also may be interested in the effect of interdependence of rating factors, assessed by the inclusion of interaction terms.

**4.1.4.1 Main Effects Fitted Together, No Interaction**

Because all three rating factors are statistically significant, they will need to be included together in a model in order to assess, as accurately as possible, the rating required for a given combination of factor levels. The simplest type of model structure accounting for all three rating factors is fitted by including the main effects without interaction terms. The GLIM notation for this model is $A+D+C$. The parametric representation of the linear predictor for cell $(i, j, k)$
where $i$ represents age at entry, $j$ represents policy duration, and $k$ represents presence of complications, is given by:

$$\beta^T z_{ijk} = b + \alpha_i + \delta_j + \gamma_k.$$  

The associated mortality ratios are found by exponentiating the linear predictor, thus:

$$e^{\beta^T z_{ijk}} = \exp(b + \alpha_i + \delta_j + \gamma_k)$$

that is, the effects are multiplicative.

The mortality ratio of 18.39 for the age at entry group 16 to 39 was based on only nine deaths. Therefore, it was decided to combine ages at entry 16 to 39 and 40 to 49 when considering more complex models, resulting in only three levels for the age at entry factor ($i = 1, 2, 3$).

The parameter estimates obtained by fitting model $A + D + C$ are as follows:

$$\hat{\beta} = 1.910, \quad \hat{\alpha}_1 = 0, \quad \hat{\alpha}_2 = -0.5668, \quad \hat{\alpha}_3 = -1.354,$$

$$\hat{\delta}_1 = 0, \quad \hat{\delta}_2 = -0.4109, \quad \hat{\delta}_3 = -0.3354, \quad \hat{\gamma}_1 = 0, \quad \hat{\gamma}_2 = 0.3359.$$  

The mortality ratios calculated for each combination of $i, j,$ and $k$ are shown in Table 3. A direct result of using the multiplicative model, without interaction terms, is that there is an underlying pattern in the mortality ratios in Table 3. Close inspection reveals that:

- Entries for "with complications" are 1.4 times larger than entries for "without complications";
- Entries in the second row are 0.66 times entries in the first row, and entries in the third row are 0.72 times entries in the first row; and
- Entries in the second column are 0.57 times entries in the first column, and entries in the third column are 0.26 times entries in the first column.

There is no conflict between the results shown here and the results for the main effects models (i.e., figures are of the same order and changes are in the same direction). The advantage is that more information is conveyed using simple mathematical relationships. Furthermore, the 18 entries in the tables of mortality ratios are derived from just six parameter estimates.
England and Haberman

TABLE 3
Impairment of Coronary Arteries: Mortality Ratios, Model A+D+C
(Multiplicative Structure)

<table>
<thead>
<tr>
<th>Duration</th>
<th>Age at Entry</th>
<th>16 to 49</th>
<th>50 to 59</th>
<th>60 to 79</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without Complications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 2</td>
<td></td>
<td>675</td>
<td>365</td>
<td>176</td>
</tr>
<tr>
<td>2 to 5</td>
<td></td>
<td>446</td>
<td>254</td>
<td>116</td>
</tr>
<tr>
<td>5 to 8</td>
<td></td>
<td>486</td>
<td>277</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>With Complications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 2</td>
<td></td>
<td>945</td>
<td>539</td>
<td>246</td>
</tr>
<tr>
<td>2 to 5</td>
<td></td>
<td>624</td>
<td>356</td>
<td>162</td>
</tr>
<tr>
<td>5 to 8</td>
<td></td>
<td>680</td>
<td>388</td>
<td>177</td>
</tr>
</tbody>
</table>

Whereas results for the main effects models fitted separately can be reproduced using traditional methods, the above results, based on the main effects fitted together, cannot be so reproduced.

4.1.4.2 Interaction Terms

The significance of interdependence between rating factors can be assessed by fitting models including interaction terms. In GLIM notation, a model includes interaction terms if an asterisk (*) appears between the symbols for model factors. For example, $A^*C+D$ represents a model including all three factors and the interaction between age at entry and the presence or absence of complications. In this example (concerning impairment of coronary arteries), the models that need to be investigated are:

- $A^*C+D$ with parametric representation
  \[ \beta^T z_{ijk} = b + \alpha_i + \delta_i + \gamma_k + (\alpha \gamma)_{ik}; \]
- $C^*D+A$ with parametric representation
  \[ \beta^T z_{ijk} = b + \alpha_i + \delta_i + \gamma_k + (\delta \gamma)_{jk}; \]
- $A^*D+C$ with parametric representation
  \[ \beta^T z_{ijk} = b + \alpha_i + \delta_i + \gamma_k + (\alpha \delta)_{ij}. \]

These models can be fitted in GLIM and the difference in deviances between model $A+D+C$ and these models referred to the appropriate
\( \chi^2 \) distribution to assess the statistical significance of the interaction terms, as shown in Table 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>Degrees of Freedom</th>
<th>Differences</th>
<th>Degrees of Freedom</th>
<th>Tail Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+D+C</td>
<td>23.647</td>
<td>18</td>
<td>1.93</td>
<td>2</td>
<td>35%</td>
</tr>
<tr>
<td>A*C+D</td>
<td>21.718</td>
<td>16</td>
<td>1.67</td>
<td>2</td>
<td>45%</td>
</tr>
<tr>
<td>C*D+A</td>
<td>21.973</td>
<td>16</td>
<td>3.60</td>
<td>4</td>
<td>47%</td>
</tr>
<tr>
<td>A*D+C</td>
<td>20.050</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results indicate that none of the first order interaction terms is statistically significant.

The only other model that can be fitted is model \( A*D*C \), the saturated model. This model reproduces the crude mortality ratios for each combination of \( i, j, \) and \( k \) and will have a deviance of zero because it gives a perfect fit, but no model simplification. The saturated model is the only other model that can be obtained from traditional actuarial methods, but it is unnecessarily complex because interaction terms are not statistically significant. This leaves the model \( A+D+C \) as the optimal model in that it is parsimonious and conveys the salient features of the data available. An examination of the associated residual plots supports this conclusion. (These plots are not shown here).

4.2 Hypertension

4.2.1 Classification

It is customary to classify hypertension as primary (essential), constituting the vast majority, or secondary to a long list of diseases (some pathological process). In the Prudential study, the hypertension group refers to primary hypertension only.

As Singer and Levinson (1976) point out, “blood pressure may be considered elevated only in terms of some normal standard.” The New York Heart Association (1955) proposes that “Any blood pressure combination up to and including 139/89 (139 mm Hg systolic and 89 mm Hg diastolic) is regarded as normotensive. Any combination including a systolic pressure of 160 and up, or a diastolic pressure of 95 and up, or both, is classified as definitely hypertensive. Any com-
Combination below 160/95 is classified as borderline hypertensive provided it is not within the normotensive limit.”

Singer and Levinson (1976) and Brackenridge (1985) report that the above definition “has been widely accepted,” and both use it in their analyses. Furthermore, it generally is accepted that blood pressure rises gradually as age increases, and increased levels in older age groups still may be compatible with average mortality. Also, significant differences in mortality with blood pressure level are observed in the normal or normotensive range.

The data were partitioned according to:

- **Age at Entry**: taking four levels as defined earlier;
- **Policy Duration**: taking six levels
  - (1) 0 to 2 years
  - (2) 2 to 5 years
  - (3) 5 to 10 years
  - (4) 10 to 15 years
  - (5) 15 to 20 years
  - (6) Over 20 years;
- **Family History**: taking two levels
  - (1) Good family history of cardiovascular disease
  - (2) Poor family history of cardiovascular disease;
- **Blood Pressure**: taking nine levels, classified simultaneously according to diastolic blood pressure taking three levels
  - (1) Under 95 mm mercury
  - (2) 95 to 105 mm mercury
  - (3) Over 105 mm mercury and systolic blood pressure taking three levels
    - (1) Under 150 mm mercury
    - (2) 150 to 165 mm mercury
    - (3) Over 165 mm mercury;
- **Weight Levels**: taking two levels
  - (1) Within 20% of standard weight
  - (2) More than 20% above standard weight; and
- **Calendar Year of Entry**: taking eight levels
  - (1) 1947 to 1951
  - (2) 1952 to 1956
  - (3) 1957 to 1961
  - (4) 1962 to 1966
  - (5) 1967 to 1971
  - (6) 1972 to 1976
  - (7) 1977 to 1981
  - (8) 1982 to 1986.

### 4.2.2 Results for Male Lives: Null and Main Effects Models

Taken as a group, the overall mortality ratio for male hypertensive is found to be 154% (based on 3,548 deaths).

We now will consider the main effects fitted separately.

#### 4.2.2.1 Age at Entry (Factor A)

The estimated mortality ratios are:
Age at Entry

<table>
<thead>
<tr>
<th></th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 to 39</td>
<td>177 (450)</td>
</tr>
<tr>
<td>40 to 49</td>
<td>210 (1029)</td>
</tr>
<tr>
<td>50 to 59</td>
<td>139 (1127)</td>
</tr>
<tr>
<td>60 to 79</td>
<td>126 (942)</td>
</tr>
</tbody>
</table>

The underlying numbers of observed deaths ($d_j$) are shown in parentheses.

Excess mortality is higher for ages at entry below age 50, as would be expected. A surprising feature here, however, is the rise in excess mortality (from +77% to +110%) for the age at entry group 40 to 49 compared with age at entry 16 to 39.

4.2.2.2 Policy Duration (Factor D)

The estimated mortality ratios are:

<table>
<thead>
<tr>
<th>Policy Duration</th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 2 years</td>
<td>155 (279)</td>
</tr>
<tr>
<td>2 to 5 years</td>
<td>135 (584)</td>
</tr>
<tr>
<td>5 to 10 years</td>
<td>164 (1120)</td>
</tr>
<tr>
<td>10 to 15 years</td>
<td>175 (765)</td>
</tr>
<tr>
<td>15 to 20 years</td>
<td>155 (443)</td>
</tr>
<tr>
<td>over 20 years</td>
<td>128 (357)</td>
</tr>
</tbody>
</table>

Excess mortality falls after the first two years duration then rises steadily to a peak at 15 years duration, after which excess mortality falls again.

4.2.2.3 Family History (Factor H)

<table>
<thead>
<tr>
<th></th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>148 (2645)</td>
</tr>
<tr>
<td>Poor</td>
<td>177 (903)</td>
</tr>
</tbody>
</table>

These results clearly show a rise in excess mortality associated with a family history of cardiovascular disease.

4.2.2.4 Blood Pressure (Factor B)

The estimated mortality ratios are:
As expected, the mortality ratios increase with increasing blood pressure (from top left to bottom right). The pressure levels shown here are consistent with the definitions of hypertension as given earlier. None of the associated mortality ratios are below 125%; therefore we clearly are dealing with blood pressure levels outside the normal (or normotensive) range.

4.2.2.5 Weight Levels (Factor W)

The estimated ratios are:

<table>
<thead>
<tr>
<th>Standard ± 19%</th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard + 20% or over</td>
<td>153 (2914)</td>
</tr>
<tr>
<td>162 (634)</td>
<td></td>
</tr>
</tbody>
</table>

Although there is a slight increase in extra mortality associated with overweight, this increase is not as large as might have been expected.

4.2.2.6 Calendar Year of Entry (Factor Y)

The estimated mortality ratios are:

<table>
<thead>
<tr>
<th>Calendar Year of Entry</th>
<th>MR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947 to 1951</td>
<td>157 (694)</td>
</tr>
<tr>
<td>1952 to 1956</td>
<td>154 (842)</td>
</tr>
<tr>
<td>1957 to 1961</td>
<td>156 (655)</td>
</tr>
<tr>
<td>1962 to 1966</td>
<td>170 (639)</td>
</tr>
<tr>
<td>1967 to 1971</td>
<td>186 (274)</td>
</tr>
<tr>
<td>1972 to 1976</td>
<td>157 (167)</td>
</tr>
<tr>
<td>1977 to 1981</td>
<td>105 (205)</td>
</tr>
<tr>
<td>1982 to 1986</td>
<td>104 (72)</td>
</tr>
</tbody>
</table>

The mortality ratios for calendar years of entry 1947 to 1961 are surprisingly stable (approximately 155%). Beyond 1961, the mortal-
ity ratios rise, reaching a peak for calendar years of entry 1967 to 1971. Beyond 1971, the mortality ratios fall until there is almost no excess mortality. These trends are difficult to interpret and may reflect changes in underwriting standards within the Prudential Assurance Company. Also, it would be problematic to extrapolate this pattern of ratios beyond 1982 to 1986.

4.2.2.7 Significance of Main Effects

Referring differences in model deviances (Table 5) to the appropriate $\chi^2$ distribution reveals that all the main effects are highly significant with the exception of weight, which is nonsignificant (although there is some evidence of a higher mortality ratio with higher weight levels). Consequently, the weight factor is dropped from subsequent model fitting.

This result that overweight in conjunction with hypertension does not add significantly to the excess mortality risk may be a surprise, but such a feature has been noted by earlier investigators; see Clarke (1961), Preston and Clarke (1966), Clarke (1979), and Leighton (1987). This may be explained by considering that the effect of an individual with hypertension also being overweight may have been allowed for in their elevated levels of blood pressure.

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>Degrees of Freedom</th>
<th>Differences</th>
<th>Degrees of Freedom</th>
<th>Tail Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>3615.7</td>
<td>3808</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3464.7</td>
<td>3805</td>
<td>151.0</td>
<td>3</td>
<td>.05%</td>
</tr>
<tr>
<td>B</td>
<td>3509.6</td>
<td>3802</td>
<td>106.1</td>
<td>6</td>
<td>.05%</td>
</tr>
<tr>
<td>D</td>
<td>3575.3</td>
<td>3803</td>
<td>40.4</td>
<td>5</td>
<td>.05%</td>
</tr>
<tr>
<td>H</td>
<td>3594.5</td>
<td>3807</td>
<td>21.2</td>
<td>1</td>
<td>.05%</td>
</tr>
<tr>
<td>W</td>
<td>3614.0</td>
<td>3807</td>
<td>1.7</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>Y</td>
<td>3553.2</td>
<td>3801</td>
<td>62.5</td>
<td>7</td>
<td>.05%</td>
</tr>
</tbody>
</table>

More complicated models (other than main effects fitted separately) may be fitted and the significance of interaction terms assessed. The results from the more important of these models will be reported. In presenting the results, it is useful to think in terms of a parametric representation of the GLIM models.
An additional parameter, $b$, is involved similar to the constant coefficient in conventional linear regression.

4.2.3 Main Effects Fitted Together, No Interaction

The GLIM notation for this model is $A + B + Y + D + H$, with parametric representation of the mortality ratio given by:

$$e^{b + \alpha_i + \pi_j + \gamma_k + \delta_l + \rho_m}$$

and, as noted before, the effects are multiplicative.

**TABLE 6**

Hypertension: Parameter Estimates for the Main Effects Model With No Interactions

<table>
<thead>
<tr>
<th>Age at Entry</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-39</td>
<td>1.00</td>
</tr>
<tr>
<td>40-49</td>
<td>1.09</td>
</tr>
<tr>
<td>50-59</td>
<td>0.70</td>
</tr>
<tr>
<td>60-79</td>
<td>0.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pressure</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systolic Pressure (mm Hg)</td>
<td>Parameter Estimates</td>
</tr>
<tr>
<td>&lt;150-</td>
<td>1.00</td>
</tr>
<tr>
<td>150-165</td>
<td>1.00</td>
</tr>
<tr>
<td>&gt;165</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Calendar Year of Entry</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>47-51</td>
<td>1.00</td>
</tr>
<tr>
<td>52-56</td>
<td>0.98</td>
</tr>
<tr>
<td>57-61</td>
<td>0.94</td>
</tr>
<tr>
<td>62-66</td>
<td>0.94</td>
</tr>
<tr>
<td>67-71</td>
<td>0.99</td>
</tr>
<tr>
<td>72-76</td>
<td>0.99</td>
</tr>
<tr>
<td>77-81</td>
<td>0.99</td>
</tr>
<tr>
<td>82-86</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>1.00</td>
</tr>
<tr>
<td>2-5</td>
<td>0.90</td>
</tr>
<tr>
<td>5-10</td>
<td>1.01</td>
</tr>
<tr>
<td>10-15</td>
<td>0.96</td>
</tr>
<tr>
<td>15-20</td>
<td>0.80</td>
</tr>
<tr>
<td>&gt;20</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Family History

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Poor</td>
</tr>
</tbody>
</table>
This model caters for all five (significant) factors simultaneously. The parameter estimates resulting from fitting this model are shown in Table 6. Mortality ratios then may be deduced from Table 6 by forming the product of relevant entries (and multiplying by 100 to express the ratio as a percentage).

Consider a hypothetical example: A man took out a whole life policy in 1977 at age 45. Upon medical examination his blood pressure was recorded as 155/100. From the proposal form it was found that his family history of cardiovascular disease could be classified as good. The policy now has been in existence for ten years, and an estimate of the excess mortality associated with this risk is required for the remainder of the policy.

<table>
<thead>
<tr>
<th>Policy Duration</th>
<th>Mortality Ratio</th>
<th>Excess Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 to 15 years</td>
<td>$1.95 \times 1.09 \times 1.18 \times 0.60 \times 1.00 \times 0.96 = 1.44$</td>
<td>+44%</td>
</tr>
<tr>
<td>15 to 20 years</td>
<td>$1.95 \times 1.09 \times 1.18 \times 0.60 \times 1.00 \times 0.80 = 1.20$</td>
<td>+20%</td>
</tr>
<tr>
<td>&gt; 20 years</td>
<td>$1.95 \times 1.09 \times 1.18 \times 0.60 \times 1.00 \times 0.72 = 1.08$</td>
<td>+8%</td>
</tr>
</tbody>
</table>

4.2.3.1 Residual Plots

If a model provides a good fit, a histogram of deviance residuals should be approximately bell-shaped (i.e., approximately normal). Also, a scatter plot of deviance residuals against linear predictor should show a corridor of values. Any other patterns would be indicative of lack of fit. (These plots are not shown.)

4.2.4 First Order Interactions

Models including first order interaction terms have been fitted and their deviances analyzed to assess the significance of the first order interaction terms. The results are not given here, but are reported in detail in England (1993). The results indicate that the interaction between blood pressure and policy duration is statistically highly significant.

5 Extensions and Further Applications

The resulting approach can be extended to incorporate different choices for the function $h()$ appearing in the relationship:

$$\theta_{\mathbf{z}} = h(\mathbf{\beta}^\top \mathbf{z}).$$
Two other choices have been explored in detail by England (1993).

\begin{align*}
\text{ADDITIVE} & \quad \theta_z = \beta^T z \\
\text{POWER} & \quad \theta_z = (\beta^T z)^{1/\gamma}
\end{align*}

The additive model is essentially the basis underlying the numerical rating system, which is widely-used for the risk evaluation of impaired insured lives as part of the underwriting process. The mortality ratio is represented by a series of parameter estimates (which may be positive or negative) that are summed. The estimates are analogous to the debits and credits used in the numerical rating system. Interdependence of the rating factors may be accommodated by including interaction terms in the definition of $z$.

The power model represents a family of models because the parameter $\gamma$ may take any real value. When $\gamma = 1$, the additive model is obtained, whereas in the limit as $\gamma \to 0$, the multiplicative model is obtained; see McCullagh and Nelder (1989). For values of $\gamma$ between 0 and 1, the power model may be regarded as being intermediate between the additive and multiplicative cases. The value of $\gamma$ giving the optimum fit, however, may lie outside the range $(0,1)$.

This modeling approach can be extended with the inclusion of approximate confidence intervals for the mortality ratios. Given the form of the mortality ratio

\[ \hat{\theta}_z = h(\beta^T z) \]

in the more general case, this procedure is not completely straightforward. England (1993) provides further details.

The approach of generalized linear modeling has been used more widely than modeling excess mortality. In particular, in the field of graduation these techniques have been used to deal with the:

- Graduation of mortality rates (Renshaw, 1991);
- Graduation in the presence of duplicate policies (Renshaw, 1992);
- Graduation of select mortality rates (Currie and Waters, 1991); and
- Graduation of transition intensities in a multiple state model (Renshaw and Haberman, 1992).
These techniques could be used for the graduation of mortality rates where it is intended to use a modification of a given standard life table.

6 Conclusions

This article attempts to highlight the benefits and power of the multiplicative hazards/generalized linear model approach. The principal advantages of this approach over traditional methods are:

- It enables comprehensive statistical analysis, including significance testing, model building, and residual analysis; and
- It allows the effect on excess mortality experience of complex interactions between the covariates to be assessed.

The approach described in this article provides a more dynamic method of constructing and testing models than the traditional approach. The current approach allows an assessment to be made of the relationship between individual factors and their interactions and their impact on excess mortality. The models of this paper do not require extensive assumptions and, with the aid of modern statistical software packages such as GLIM, can be implemented easily. As we have noted, these models can be seen to be a direct generalization of traditional actuarial mortality ratios.

References


England and Haberman

Excess Mortality


Appendix—The Aggregate Integrated Standard Force of Mortality

Recall the definition of $e_j$ given in equation (7), i.e.,

$$e_j = \sum_{k=1}^{N_j} \int_{t_{jk}}^{T_j} \mu^*(u) \, du.$$  \hfill (7)

We will now show that $e_j$ can be interpreted as the expected number of deaths in cohort $j$ had standard mortality rates applied. This interpretation can be justified because the expected value of $e_j$ is equal to the expected number of deaths had standard mortality rates applied. This is proved briefly by Berry (1983). A more complete proof is shown below.

Consider a complete follow-up study, i.e., one where there are no withdrawals or losses. This assumption is being made to simplify the presentation. For cohort $j$, let us assume that individual $k$ ($k = 1, 2, \ldots, N_j$):

- Enters the study at time 0 (so $\tau_{jk} = 0$). This can be done by a simple change of origin;
- Has a maximum follow-up time of $T_{jk}$;
- Exits the study at time $T_{jk}^*$ (which is $t_{jk}$ in equation (7)); and
- Has an indicator random variable $I_{jk}$, where $I_{jk} = 0$ if individual $k$ leaves the study alive ($T_{jk}^* > T_{jk}$) and $I_{jk} = 1$ if individual $k$ dies during the study ($T_{jk}^* \leq T_{jk}$).

The probability distribution function for $T_{jk}^*$ is $f^*(t)$ (assuming that standard mortality applies) and $F^*(t)$ is its cumulative distribution function. The force of mortality is $\mu^*(t)$ where

$$\mu^*(t) = \frac{f^*(t)}{1 - F^*(t)}.$$  \hfill (A1)

Let

$$e_{jk} = \int_0^{T_{jk}^*} \mu^*(t) \, dt.$$  

Clearly $e_{jk}$ is a random variable because $T_{jk}^*$ is a random variable. To calculate the expected value of $e_{jk}$, it is necessary to consider the possibility of death before $T_{jk}$ or after $T_{jk}$. This leads to:
The first part of equation (A1) relates to the contribution to the expected value made by the possibility of death occurring at time \( s \), integrated over all possible values of \( s \) from 0 to \( T_{jk} \). The second part is the contribution made by the possibility of survival to time \( T_{jk}' \), the maximum follow-up time.

Integrating the first component of equation (A1) by parts gives:

\[
E[e_{jk}] = \int_0^{T_{jk}} f^*(s) \int_0^s \mu^*(t) \, dt \, ds + (1 - F^*(T_{jk})) \int_0^{T_{jk}} \mu^*(t) \, dt.
\]  
(A2)

The right side of equation (A3) follows from equation (A1). Next we will prove that \( E[e_{jk}] = E[I_{jk}] \). The contribution that individual \( k \) makes to the number of deaths is 0 if the individual survives to \( T_{jk} \) (with probability \( 1 - F^*(T_{jk}) \)), and it is 1 if the individual dies before \( T_{jk} \) (with probability \( F^*(T_{jk}) \)). Hence,

\[
E[I_{jk}] = 0 \times \text{Prob(survival to } T_{jk}) + 1 \times \text{Prob(death before } T_{jk})
\]

\[
= 0 \times (1 - F^*(T_{jk})) + 1 \times F^*(T_{jk}) = F^*(T_{jk}).
\]

Therefore \( E[e_{jk}] = E[I_{jk}] \).

Now for the entire cohort \( j \), the term \( e_j \) is defined as

\[
e_j = \sum_{k=1}^{N_j} e_{jk}
\]  
(A4)
and the expected number of deaths (had standard mortality applied in this cohort), \( d^*_j \), is given by

\[
d^*_j = \sum_{k=1}^{N_j} E[I_{jk}] = E[e_j]. \tag{A5}
\]

Equation (A5) shows that \( e_j \) is an unbiased estimate of \( d^*_j \). Therefore, the statement that the aggregate integrated standard force of mortality in cohort \( j \) can be interpreted as the expected number of deaths in cohort \( j \) had standard mortality rates is justified.

Peter D. England  
Commercial Union Assurance plc  
6 Broadgate (Level 7)  
London, England EC2M 2QS

Steven Haberman  
Department of Actuarial Science and Statistics  
The City University  
Northampton Square  
London, England EC1V 0HB
The Small Plan Audit Program: The Opinions of the Court

Arnold F. Shapiro

Abstract

One of the most important issues of recent years from the perspective of many pension actuaries is the IRS's small plan audit program. The program initially was expected to raise two-thirds of a billion dollars by targeting well-funded defined benefit plans with five or fewer participants. The focus of the audit was the assumed interest rate and the normal retirement age, both of which the IRS generally regarded as too low.

While the focus of the audit was relatively narrow, the issue it raised was a fundamental one. The basic question was the extent to which the IRS could impose its unilateral interpretation of actuarial principles on pension actuaries.

Not surprising, many small plan audit cases ended in the tax courts. In due course decisions and opinions have been rendered in three lead cases. This article presents the opinions of these cases as they relate to actuarial practice and discusses some of their implications.

Key words and phrases: defined benefit plans, actuarial assumptions, unit credit method, IRS

1 Introduction

One of the most important issues of recent years from the perspective of many pension actuaries is the Internal Revenue Service's (IRS) small plan audit program. The program began in November 1989, when the IRS initiated a nationwide plan to audit the actuarial assumptions of approximately 18,000 small well-funded defined benefit plans. The program began in November 1989, when the IRS initiated a nationwide plan to audit the actuarial assumptions of approximately 18,000 small well-funded defined benefit plans. The program began in November 1989, when the IRS initiated a nationwide plan to audit the actuarial assumptions of approximately 18,000 small well-funded defined benefit plans.

1 Even though this paper deals only with court cases in the United States, the opinion of the court may have implications in any country where actuarial assumptions are at issue.

2 Throughout this paper, the abbreviation IRC means the Internal Revenue Code and the abbreviation IRS refers to the Internal Revenue Service of the U.S.
fit pension plans. The program initially was expected to raise two-thirds of a billion dollars in additional tax revenue.

The program appeared to be floundering; see the BNA Pension Reporter (1992). In retrospect this is not surprising because the effort immediately met intense and unrelenting resistance from small plan actuaries, their associations, and their advocates.

It was not long after the small plan audit program was instituted before several of the ensuing cases reached the tax court. These cases were assigned to Judge Charles E. Clapp II, who, after observing that there were likely to be many more such cases, selected some representative ones for trial. His stated intent was to develop judicial precedence and guidance so that subsequent cases could be resolved without costly litigation.

The suits comprise two institutional and eight noninstitutional cases. The two institutional cases, the first to be tried, involved large successful law firm partnerships that had adopted individual defined benefit (IDB) plans for their partners. The Texas-based firm of Vinson & Elkins and the New York firm of Wachtell, Lipton, Rosen & Katz (Wachtell Lipton). In both instances, assumptions used for valuing their plans were deemed

3 These plan years were chosen because the statute of limitations was ended for plan years prior to 1986 (IRC §6501) and the tax law changed for plan years that ended after 1988. The primary relevant changes in the tax law were the revision of the full-funding limitation to include current liability (IRC §412(b)(5), (c)(7) and (l)(7)) and the amendment to IRC §412(c)(3), which requires that each actuarial assumption (rather than actuarial assumptions in the aggregate) be reasonable.

4 In view of IRC §401(a)(26), individual defined benefit plans of this type no longer are allowed, and these plans have been terminated.
unreasonable by the IRS, which sought to disallow their deductions. These cases were tried in January 1992, and a decision was handed down the following July.

The remainder of the cases involved a variety of small businesses, each of which had a small defined benefit pension plan for one or two key employees. Because the cases arose under an audit program in Phoenix, they came to be known as the Phoenix cases, but subsequently were referred to as Citrus Valley because they were consolidated and tried as *Citrus Valley Estates, Inc. et al.* These cases involved frontloading of the contribution under the unit credit funding method in addition to actuarial assumption challenges. The cases were tried in February 1992, and a decision was handed down the following September.

This article presents the opinions of the court as they relate to the actuarial practice associated with small defined benefit plans and discusses some of their implications. First, the actuarial issues contested by the IRS are summarized. Then the opinions of the court relating to these issues are discussed. The paper ends with a comment on the implications of the court's opinions.

2 The Actuarial Issues Contested by the IRS

The general actuarial issue raised by the IRS was whether actuarial assumptions used by the enrolled actuary to determine the plans' costs were reasonable in the aggregate and represented the actuary's best estimate of anticipated experience under the plans as required by IRC §412(c)(3). The specific issues contested by the IRS are summarized in Table 1. For example, for the Vinson & Elkins plans the IRS contested the 5 percent preretirement and post-retirement interest rate assumption, the normal retirement age of 62, the 5 percent postretirement expense load, and the preretirement mortality assumption. Moreover, the IRS contended that these assumptions were not offset by any other assumptions that would make the assumptions reasonable in the aggregate.

---


6 This paper does not deal with the nonactuarial issues of these cases, which included timing of amendments, automatic approval of a cost method change, and validation of hours worked.
Most of the issues of Table 1 are self-evident, but those related to the mortality tables and the cost methods need clarification. For the institutional cases, the IDB plans that contained life insurance used the 1958 Commissioners Standard Ordinary (CSO) mortality table for the preretirement mortality assumption and the 1971 Individual Annuity Mortality (IAM) table for the postretirement mortality assumption. While the IRS agreed that such plans may provide a preretirement death benefit and may fund these benefits using envelope funding, it contested the use of the 1958 CSO table on the grounds that it grossly overstated the expected actual mortality experience.

TABLE 1
Actuarial Issues Contested by the IRS

<table>
<thead>
<tr>
<th>Interest Rate</th>
<th>Expenses</th>
<th>Mortality Table</th>
<th>Cost Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>Post</td>
<td>NRA</td>
<td>Pre</td>
</tr>
<tr>
<td><strong>Vinson &amp; Elkins</strong></td>
<td>5%</td>
<td>5%</td>
<td>62</td>
</tr>
<tr>
<td><strong>Wachtell Lipton</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Citrus Valley et al.</strong></td>
<td>6%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Citrus Valley</strong></td>
<td>5%</td>
<td>5%</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td><strong>Davis</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Old Frontier</strong></td>
<td>5%</td>
<td>5%</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td><strong>Lear Eye Clinic</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Robert Stephan</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Boren Steel</strong></td>
<td>5%</td>
<td>5%</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td><strong>Arizona Orthopedic</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Fox</strong></td>
<td>5%</td>
<td>5%</td>
<td>55</td>
</tr>
<tr>
<td><strong>Brody Enterprises</strong></td>
<td>5%</td>
<td>5%</td>
<td>6/4.5%</td>
</tr>
</tbody>
</table>

7 Some of these plans could be differentiated only on the basis of their credible experience. It had been anticipated that the court's decision would be affected materially by plan experience, but this turned out not to be the case.

8 The envelope method may be used with any cost method and with any type of insurance policy. It is the method that generally is used with the unit credit funding method or with insurance policies that do not have guaranteed projected cash values at retirement. Under the envelope method, assets are adjusted by adding the cash value of the insurance as of the valuation date. The normal cost and accrued liability are calculated using the adjusted assets.
The situation in Citrus Valley was somewhat different. In one instance, an insurance company’s guaranteed female annuity table was used for a plan with a single male participant; in another, a female mortality table with a seven year age setback was used for a plan with a single male participant; and in another, an age setback was used for a participant with a substandard family medical history. The IRS contested the mortality assumption in each instance.

The IRS contested the actuarial cost method in a significant number of the Citrus Valley cases. The issue was straightforward. These plans provided for the accrual of all, or a significant portion, of the benefits provided under the plan in a very few years, a procedure commonly referred to as frontloading. Using the unit credit funding method, the benefits then were funded as they accrued with the contribution currently deductible. The IRS contended that while frontloading of benefit accruals is permissible from a qualification standpoint, an equivalent frontloading of the deductible contribution is not permitted and that no more than 10 percent of the maximum benefit may be allocated to a given year’s normal cost, just as the maximum benefit that can be provided to a participant with one year of service is 10 percent of the overall IRC §415 limit.

3 The Experts

Before proceeding to the findings of the court, it is worth noting the credentials of the experts chosen by each side and the focus of their testimony or report.

3.1 The Institutional Cases

The experts for institutional cases included James F. Rabenhorst, managing partner at Price Waterhouse, who testified regarding the retirement age assumption; Richard R. Joss, Ph.D., F.S.A., M.A.A.A, E.A., resource actuary for the Wyatt Company, who testified regarding the actuarial assumptions; Mary S. Riebold, F.S.A., M.A.A.A., E.A., F.C.A., managing director for Mercer and then-president of the Conference of Consulting Actuaries, who testified regarding the actuarial assumptions; Steven H. Schechter, director of management information systems at Wolper Ross, who testified regarding interest rate assumptions based on an analysis of Form 5500 data; and John W. Peavy III, Ph.D., C.F.A., professor of finance at Southern Methodist University, who served to rebut the contentions of Shapiro and Haneberg regarding the interest rate assumption.
The experts for the IRS in these cases included Ronald L. Haneberg, J.D., F.S.A., M.A.A.A., F.C.A., previously a consulting actuary with Buck Consultants, who testified regarding the actuarial assumptions; Claude Poulin, F.S.A., M.A.A.A., E.A., president of Poulin Associates, Inc., who testified regarding the actuarial assumptions; Alan C. Shapiro, Ph.D., professor of banking and finance at the University of Southern California, who testified regarding the interest rate; William S. Borden, Ph.D., senior program analyst at Mathematica Policy Research, who testified regarding the investment return and in rebuttal to Joss; and Jeffrey F. Jaffe, M.B.A., Ph.D., associate professor of finance at the Wharton School, who provided an expert report on the validity of the interest rate assumption.

3.2 Citrus Valley Estates, Inc. et al.

The experts for Citrus Valley included Kenneth D. Klingler, F.S.A., M.A.A.A., E.A., a consulting actuary with the Wyatt Company, who testified regarding the assumptions; and Arthur W. Anderson, A.S.A., M.A.A.A., E.A., who previously had been a consultant for William M. Mercer, Johnson & Higgins, and the Wyatt Company and was the author of Pension Mathematics for Actuaries, testified as an expert with respect to the unit credit funding method.

The experts for the IRS included J. Ruben Rigel, J.D., F.S.A., F.C.A., M.A.A.A., E.A., who testified with respect to the assumptions and the unit credit funding method; Roger Ibbotson, M.B.A., Ph.D., president and chief executive officer of Ibbotson & Associates, Inc., who testified with respect to the interest rate assumption; William S. Borden, Ph.D., who testified with respect to the interest rate and retirement age assumption; and James E. Holland, A.S.A., E.A., chief of the Pension Actuarial Branch of the Service, who provided an expert report dealing with the unit credit funding method.

4 The Findings of the Tax Court

The court generally found against the IRS on most of the issues. In the institutional cases, for example, the court held that "[t]he actuarial assumptions made by the plans' enrolled actuary were reasonable in the aggregate and represented the actuary's best estimate of anticipated experience under the plans, as required by §412(c)(3); accordingly, as the assumptions used were not substantially unreasonable, [the IRS] is precluded from requiring a retroactive change of assumptions."
The court held similarly for the noninstitutional cases that all of the challenged actuarial assumptions for each of the plans at issue were reasonable. Further, the certifying actuaries for the plans using the unit credit funding method funded within allowable limits and made reasonable allocations of costs, except for one plan that was complicated because of an amendment issue (Citrus Valley, p. 101). Accordingly, the actuarial assumptions and methods used for the plans were reasonable in the aggregate. A fortiori, these assumptions were not substantially unreasonable in order to permit retroactive changes of assumptions for years prior to the year in which the audit was made.

The outcomes of the cases were not obvious prior to the decisions. It is interesting and informative to read how an unbiased legal authority interprets the actuarial issues involved. The following is a recapitulation of how the court reached its conclusions.

4.1 Deference to the Enrolled Actuary

A major conclusion was that deference must be given to the assumptions chosen by the enrolled actuary who certifies the funding of the plan. In this regard, Judge Clapp gave his interpretation of Congressional intent the full weight of legal authority.

Judge Clapp emphasized that Congress was aware in enacting ERISA that actuaries would play a major role in ensuring that retirement plans would be sufficiently able to provide retirement income when due. He observed that Congress recognized the importance of the actuarial assumptions and the cost methods chosen by actuaries in determining plan funding amounts and that Congress explicitly noted that such determinations by actuaries would involve making predictions and would be a matter of judgment involving many factors and producing a range of results. He also commented that Congress decided that accepting a range of reasonableness for funding amounts for retirement plans would be more desirable and more effective than imposing an inflexible legislative standard on actuaries and, therefore, rejected imposing mandatory funding assumptions and methods (Wachtell Lipton, pp. 10-11).

4.2 The Interest Rate Assumption

In reaching his decision on the interest rate assumption, Judge Clapp identified what he regarded as particularly important factors. He noted that the combination of these factors weighed heavily
in favor of concluding that 5 percent was reasonable. For the institutional plans these factors were (Vinson & Elkins, p. 46):

- The nature of the responsibility Congress entrusted to enrolled actuaries in the statutory scheme enacted for defined benefit pension plans;
- The conservative nature of the actuarial assumption selection process;
- The fact that IDB plans were long-term plans, with funding to occur over a 30 year to 50 year period;
- The fact that IDB plans were self-directed, with each participant being a co-administrator, especially because most of the IDB plans did not employ a professional manager;
- The fact that IDB plans lacked credible experience with respect to earnings, investment strategies, and otherwise;
- The risk of losing compounded earnings in a tax-exempt trust associated with using overly optimistic assumptions and the resulting requirement for unanticipated higher contributions in later years;
- The relative closeness of all the actuarial experts' reasonable ranges; and
- The fact that most actuaries used interest rate assumptions of between 5 percent and 6 percent for small plans during the years at issue.

He listed the same factors, except for the relative closeness of the reasonable ranges, for the noninstitutional plans (Citrus Valley, p. 69).

Judge Clapp also clarified the role of a prudent actuary in the selection of the interest assumption. He noted that the actuary's primary duty to plan participants under ERISA is to establish a realistic contribution pattern over the long term so that the plan sponsor will provide adequate funding for the ultimate pension obligation. Thus, prudent actuaries maintain a long-term conservative view that

---

9 It is relevant that each partner/participant served as a coadministrator because that meant that the plan assets of the IDB plans were not commingled for the purpose of investment and, therefore, could not realize the rates of return earned by larger plans.

10 Not all the experts agreed that their reasonable ranges were close. See, for example, Ronald L. Haneberg, "Not All Experts Agree," Enrolled Actuaries Report (November, 1992), p. 3.

11 This conclusion follows from Schechter's testimony that actuaries established interest rate assumptions between 5 percent and 6 percent for 1986 plans with fewer than 100 participants for 76.6 percent of the preretirement assumptions and 82.5 percent of the postretirement assumptions. Schechter's conclusions were based on his analysis of data obtained from the Department of Labor.

The court was not swayed by the IRS's contention that rates in general use during the time were irrelevant.
will ensure benefit security for plan participants in selecting actuarial assumptions (Vinson & Elkins, p. 27)

Rejecting the IRS’s contention that 8 percent would have been a reasonable interest rate assumption because that amount could have been earned during the years at issue, the court commented that “Congress did not entrust the nation’s tax-advantaged retirement savings system to hypothetical returns that the markets ‘should’ bear” (Vinson & Elkins, p. 49).

Particularly noteworthy is the fact that the court attached only minor importance to the testimony and reports of nonactuaries, in spite of the fact that they were experts in the field of investment. This was true, for example, even with the testimony of the well-known Roger Ibbotson.12 The rationale was that these persons were not actuaries and that conclusions they drew would have limited application in the determination of the reasonableness of actuarial assumptions (Vinson & Elkins, p. 47). The court reasoned that if a financial analyst’s predicted rate is higher than the actual rate earned, the investor simply earns less than expected, whereas if an actuary makes the same mistake, there is a significant risk that the plan will become underfunded and the pensioners’ full benefits will be unpaid (Citrus Valley, p. 71).

4.3 Retirement Age Assumption

The court seemed willing to accept a normal retirement age (NRA) assumption that was less than age 65 as long as it was based on reasons that were “sincere, credible, and reasonable.” It explicitly rejected the IRS’s argument that statements by the participant in a one person plan were merely self-serving, even when there was no evidence that the underlying reasons had been explained to the plan actuary. (See, for example, Citrus Valley, p. 83.)

The IRS took the position that failure of a key participant to retire at the assumed normal retirement age was clear evidence that the assumption was unreasonable. In rejecting this position, the court noted that “... the certifying actuary is not charged with the responsibility of determining when a plan participant will actually begin to receive the plan benefits. That would be an impossible task. Further, the fact that a plan participant might choose to, or actually does, delay receipt of the plan benefits beyond the assumed retire-

12 Ibbotson & Associates, Inc. sells financial software and data and provides consulting services to investment management firms. Roger Ibbotson is an often-quoted authority on stocks, bonds, Treasury bills, and inflation.
ment age does not make the retirement age assumption unreasonable. An actuary is charged with looking into the future and making a determination as to, among other things, when benefits under the plan could begin" (Vinson & Elkins, p. 58).

Some of the Citrus Valley plans contained a segregated account provision, which meant that at the normal retirement age benefits were segregated into a separate account even if the participant chose to continue working beyond that age. The present value of accrued benefits at the normal retirement age is treated in effect as an account balance in a defined contribution plan. The experts of both parties agreed that the inclusion of a segregated account provision in a plan rendered the date of a participant’s actual retirement irrelevant (Citrus Valley, p. 75).

Given that the experts agreed, the court concluded that the segregation provision justified the finding that it was reasonable for the assumed retirement age to be the normal retirement age stated in the plan, because that would be the age at which the participant would elect to segregate the accrued benefits. This obviated the retirement issue for a number of plans that had a normal retirement age of 55.

4.4 Expense Loadings

The court held for the taxpayer in each instance where the IRS challenged the expense loading. While Judge Clapp had some misgivings about the 7.5 percent expense loading in the institutional cases, he found it not to be substantially unreasonable and acceptable on the basis of reasonable in the aggregate.

He rejected the IRS’s argument in the noninstitutional cases that expense loading is merely a device to increase deductions. His opinion observed that “[the IRS] offered a rather perfunctory rebuttal, stating simply that [the] addition of postretirement expense load assumptions would further increase the funding goal and the amount of the deduction ... This is not, however, unreasonable per se, as [the IRS] seems to believe ... A postretirement expense load is a reasonable manner in which to fund the postretirement administrative fees” (Citrus Valley, p. 91).

4.5 Mortality Assumptions

The court found that it was reasonable in the institutional cases to use the 1958 CSO mortality table to compute the cost of the pre-retirement death benefit. It explicitly rejected the IRS’s arguments that a preretirement mortality assumption was unreasonable in a one
person plan and that even if it were appropriate to use a preretirement mortality assumption, it was unreasonable to assume the 1958 CSO mortality table for the preretirement mortality and the 1971 IAM table for the postretirement mortality for the same person because the tables are incompatible. As the court pointed out, the probability of the participant’s preretirement death was not at issue. The issue was to estimate the life insurance premium expense, and this could be done best by using the same type of mortality table as would be used by the insurance company (Vinson & Elkins, p. 67).

In the noninstitutional cases, while the court was “not entirely convinced that the mortality assumption ... is completely reasonable, it is not substantially unreasonable so as to justify a retroactive adjustment” (Citrus Valley, p. 87). Thus, even in situations as extreme as the case involving a male participant that used the 1983 IAM table for females with a seven year age setback, the mortality assumption implicitly was approved by the court in its approval of the funding assumptions in the aggregate.

4.6 The Unit Credit Funding Method

One of the surprises to emerge from the Citrus Valley cases was the court finding against the IRS on the frontloading issue under the unit credit funding method. The IRS previously had won the well-publicized Mirza case (Jerome Mirza & Associates, Ltd. v. United States, 882 F.2d 229 (7th Cir. 1989)), where the same issue was in question and the same argument was used. In Mirza, the court agreed with the IRS’s interpretation that IRC §404(a)(1)(A)(iii) provides that the maximum that can be deducted in any year is the “normal cost” plus an amount necessary to amortize “past service” and other supplementary cost over ten years, as determined under regulations prescribed by the Secretary. It reasoned that “[i]t is simply inconceivable that Congress would take pains to provide for the amortization of past service credits but intended to allow taxpayers to circumvent this requirement by the device of structuring their plans to accrue benefits in a single year” (Mirza, p. 232)

Judge Clapp enumerated three reasons for rejecting the Mirza conclusion (Citrus Valley, pp. 104-105). First, “[t]he language of §404(a)(1)(A)(iii) setting forth the limit on deductible contributions used the conditional phrase ‘if *** provided by the plan’ when setting forth the treatment for past service cost.” Thus, there would be only a past service liability if it were provided by the plan. Second, “[d]espite [the IRS’s] assertions to the contrary, there is no express[ed] or implied connection between the limitations of §415 and
any allocation under §1.412(c)(3)-1(e)(3)” (Citrus Valley, p. 99). That is, there is no requirement that the allocation between normal cost and past service liability be consistent with the limitations on benefit accruals. Third, “the Unit Credit Funding Method—in connection with a career-average pay plan—inherently allocates benefits in a reasonable manner to the past and future years of service for which benefits accrued and will accrue.”

This finding is only relevant for plan years beginning prior to 1987, as the approach discussed is not possible for plan years beginning after December 31, 1986. The Tax Reform Act of 1986 amended §415(b)(5) so that the dollar limitation is phased in over the first ten years of participation in a plan rather than ten years of service with the plan sponsor.

4.7 Evidentiary Matters

The IRS consistently has objected to actuaries’ use of its training manuals, audit guidelines, internal and external correspondence, and transcripts of speeches made by Service employees regarding the matters at issue in these cases. The court concluded (Vinson & Elkins, pp. 75-77), however, that actuaries can take into account IRS documents that have been disseminated publicly because “they are part of the actuarial universe within which all actuaries must live, think, and work in arriving at their conclusions as to reasonableness and their best estimates regarding appropriate contributions.” Moreover, actuaries can be guided by the speeches of high-ranking Service employees.

One specific comment that had been referenced by many pension actuaries is the highly publicized transcript of the Ira Cohen speech at the 1986 Enrolled Actuaries meeting, wherein he stated that a 4 percentage point corridor on either side of the prevailing long-term Treasury bond rate was within the reasonable range of interest rate assumptions. In spite of the fact that Cohen was the director of the Actuarial and Technical Division of the Service at the time of the speech, the IRS claimed that he had not spoken for the Service and, moreover, the speech was merely hearsay. The court disagreed with the IRS’s position, and asserted that such a speech, heard by many actuaries and disseminated by publication to many more, is not hearsay, as long as the transcript is “true and correct.”
5 Implications

There seems to be a consensus among small plan attorneys that the opinions rendered in these cases are likely to be afforded considerable credibility. Not only are they "lengthy, studious, and thoroughly analyzed," but they are based to a large extent on "factual conclusions," which makes them difficult to overturn; see Reish and Ashton (1992). Moreover, 14 of 15 participating tax court judges in the Phoenix cases concurred with the opinions.

It is difficult to anticipate how the courts will react in future cases where the issues are similar, but the facts and circumstances are materially different. The following basic principles, however, seem to have emerged:

- The intent of Congress is that deference should be given to the assumptions chosen by the enrolled actuary;
- While assumptions are required to be reasonable and Congress did not permit actuaries unfettered liberty, the pragmatic test is that assumptions are not "substantially unreasonable;" and
- When formulating assumptions, it is appropriate for the actuary to be guided by the "sincere, credible, and reasonable" expectations of the plan sponsor and IRS documents and insights that have been publicly disseminated.

In the past actuaries have struggled to formulate a workable interpretation of pension laws and regulations for small plans. In most cases, actuaries are not attorneys, however—while their interpretation of these laws and regulations may have seemed reasonable to them, there has been a need for an authoritative unbiased interpretation. These cases, with their scholarly exposition of the rules and regulations, have done much to help put things into perspective.

References


14 The court specifically noted that it was the intent of Congress that actuaries should not sell their expertise to achieve tax-desired results rather than prudent plan funding.

15 It is worth noting that the two main assumptions considered (the interest rate and the retirement age) were argued successfully on an individual basis, rather than an aggregate basis, so that the conclusions reached are still appropriate.
Arnold F. Shapiro
Small Plan Audit Program

Citrus Valley Estates, Inc. et al. v. Commissioner, 99T.C. No. 21, No. 12900-89 etc., September 29, 1992


Internal memorandum to IRS field agents dated November 29, 1989.


Jerome Mirza & Associates, Ltd. v. United States, 882 F.2d 229 (7th Cir. 1989).


Arnold F. Shapiro
Penn State University
409 Business Administration Building
University Park, PA 16802
Book Reviews


Reviewer: A. Hoque Sharif

Option pricing has been an active area of research in the field of finance during the past 20 years. There are several textbooks that cover options pricing theory in great detail (including texts by Cox and Rubenstein (1985), Bookstaber (1991), Hull (1993), Jarrow and Rudd (1983), and Ritchken (1987)). Do we need another textbook? Yes, for actuaries there is a need for a book that explains the management of financial risk at an introductory level. There is no similar text written for actuaries.

In recent years, actuaries have become more involved in various aspects of corporate finance. In fact, the Society of Actuaries has already incorporated some courses on investment and finance in its education program and has opened a finance track (fall of 1993) for its Fellowship examination process. One expects Professor Boyle’s Options and the Management of Financial Risk to play an important role in this process and in the education of a new generation of actuaries.

There are two major topics covered in this text: (1) models of the term structure of interest rates, and (2) the analysis and valuation of derivative securities. Only a few basic concepts in finance are introduced, and they are dealt with at an introductory level with numerical examples. Unfortunately, Professor Boyle does not refer the reader to any introductory level text. The only other textbooks referenced are by Malliaris (1982), Merton (1990), and Van Horne (1970). None of these can be considered as an introductory text.

* A. Hoque Sharif ASA (1984) is a Ph.D. student in actuarial science in the Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada. He received his undergraduate education at the Dhaka University, Bangladesh, M.Sc. (Statistics) in 1979 from the University of Saskatchewan, and M.Math (Actuarial Science) in 1988 from the University of Waterloo. Prior to enrolling at the University of Waterloo, he was employed by the Dhaka University, the University of Saskatchewan, Reed Stenhouse Associates, Sovereign Life Insurance Company, and Wright Mogg & Associates. His current research interest is in multistate life table modeling and its applications.
Figure 1
Term Structure

Figure 2
Value of Assets and Liabilities
should be pointed out that the actual reference to Malliaris should be Malliaris and Brock (1982). In addition, a more recent edition to Van Horne’s text was published in 1990.

My only criticism of *Options and the Management of Financial Risk* is that it lacks exercises, graphs, and a subject index. None of the chapters has any exercises for readers. This is inconsistent with the philosophy of learning in the mathematical sciences, where doing actual problems is of vital importance. I hope that Professor Boyle corrects this by developing a companion text consisting of worked examples and exercises. Graphs are an efficient means of quickly communicating information, but they are used infrequently in this book. For example, inclusion of the two figures above (drawn using tables 2.2 and 2.6 on pages 28 and 35, respectively, of Professor Boyle’s text) would have conveyed the essential information at a glance and would have helped to buttress his excellent points. There are several other places where graphs would have expedited communication with the reader. A subject index would have benefited readers.

*Options and the Management of Financial Risk* covers a sufficiently broad range of topics to provide a sound introduction to the management of financial risks. The book is well written and can be covered easily in a one semester university or college course. An elementary knowledge of interest theory and probability theory is sufficient background for understanding the material presented.

The book consists of eight chapters. Chapter 1 provides an introduction to the subjects of insurable risk, financial securities, financial risk, and financial risk management. It also provides an overview of the text. Chapter 2 introduces the framework for analyzing the term structure of interest rates in a deterministic setting. Classical definitions of duration and convexity are covered in Chapter 3. Derivative securities (options, forwards, futures, and options futures) are discussed at the grassroots level in Chapter 4. In Chapter 5 several relationships that option prices must obey (namely, put-call parity) are derived using the no-arbitrage principle. Chapter 6 assembles several results from probability and statistics, including the central limit theorem, normal and lognormal distribution, and a simple random walk model, all of which are useful in option pricing. The famous Black-Scholes formula for pricing European call options is analyzed in great detail in Chapter 7. The concluding chapter deals with stochastic interest rate models and their applications.

There are several existing textbooks on option pricing, but Professor Boyle’s *Options and the Management of Finance Risk* provides an excellent starting point for actuaries, especially those unfamiliar
with modern finance theory. This book will be welcomed by actuaries.

References


A. Hoque Sharif
Department of Statistics & Actuarial Science
University of Waterloo
Waterloo ON N2L 3G1
Canada

Reviewer: Charles Fuhrer*

This 784 page book includes 34 chapters by different authors. The chapters are assigned to eight sections, each with a section editor. The principal editor is William F. Bluhm. I will comment about the book as a whole and then cover each section briefly.

The book is comprehensive and well-written. The authors and editors are to be commended for the high quality of the text. This is particularly impressive, given the difficulties associated with preparing such a large volume with so many different contributors.

The dust cover states that "Group Insurance is intended to serve as both an educational text for beginners in the field, and as a reference text for experienced practitioners." It is difficult for any text, however, to serve both beginners and experienced practitioners well. Group Insurance is an excellent text for beginners. It covers a vast amount of material (most of it in summary form) at a level appropriate for beginners. Unfortunately, this may detract from its usefulness to experienced practitioners who usually need detailed information. It is interesting to compare this book with Group Insurance Handbook (1965) which could be considered its precursor. Group Insurance Handbook contains much more detail than does Group Insurance.

Group Insurance has very few references to other articles and books. There are only 30 endnotes: 12 cite court cases or government rulings, eight are tables, two are current pamphlets, one is a current events bulletin, one is an investment-ratings publication, two are accounting standards opinions, and only four are original articles. Even the venerable Group Insurance Handbook is not mentioned anywhere.

This lack of references is unfortunate for several reasons:

- References would allow the reader to verify the accuracy of the facts that are used. Group Insurance includes many facts and figures without mentioning their sources;
- References enhance an educational work because they give the beginner (and even the experienced practitioner) a guide for further study; and

* Charles S. Fuhrer, FSA (1977) of the Washington National Insurance Co. in Lincolnshire, IL has been a group insurance actuary since 1973. He is co-editor of Actuarial Research Clearing House and has given numerous presentations at actuarial meetings. Mr. Fuhrer has written many papers and has been awarded the 1988 Practitioner's Prize by the AERF and the 1991 Health Section Research Papers Prize.
• Discussion of other works gives the reader a sense of the historical development of the subject matter and how the current material fits with other thoughts about the subject. For example, the reader will be able to determine if the current work is consistent with standard or classical thinking in a particular discipline or if the current work is new and original.

Chapter 25, “Bayesian Statistics and Credibility” by Thomas N. Herzog, is a notable exception. This chapter contains 24 endnotes, with other works mentioned throughout the chapter (particularly on page 516). The reader can understand how the author’s thoughts fit into the total work on the subject.

Comments about specific sections follow. Also included are references for some of the subjects.

• Section 1—“Introduction,” Robert B Cumming, editor. For a more detailed history of group insurance through 1965, see “Development and Significance of Group Life Insurance” by C. Manton Eddy and “Development and Significance of Group Health Insurance” by J.F. Follmann, Jr. in Group Insurance Handbook (1965). Several statements by Richard S. Bilisoly in the current Chapter 1 are very similar to those made by Eddy and by Follmann in Group Insurance Handbook.

In Chapter 2, David F. Ogden tells us what the 1990 market shares of the players are. The author should state the sources of this information.


In Chapter 5, “Group Disability Income Benefits,” David W. Simbro does not mention that under most long-term disability (LTD) plans the benefit is not reduced further by any Social Security cost of living increases that occur after the disabled individual becomes eligible to receive benefits.

Chapter 6, “Medical Benefits in the United States,” by Darrell D. Knapp, defines medical benefits by the dimensions of services and conditions, the degree to which the insured shares in the cost, and the degree to which the provider participates in the cost. This is an original and clever way of organizing a complicated subject. For historical purposes (and because many still exist today), base plus supplementary (or superimposed) major medical and comprehensive major medical plans should be defined. See the Group Insurance Handbook (1965), Chapter 18.
Chapter 7, by Bruno Gagnon, examines medical benefits in Canada, while David R. Nelson analyzes miscellaneous coverages in Chapter 8.

In Chapter 9, Bruce D. Schobel studies government plans in the United States. Robert J. Myers' Social Security (1985 and previous editions) generally has been considered to be the Bible of Social Security for actuaries. Of course, there have been many other works and numerous U.S. government publications on social insurance.

- **Section 3—"The Legal and Regulatory Environment,"** Charlotte A. Furman, editor. This section contains only a brief discussion of the legal environment; the regulatory environment is covered in much greater detail. See, for example, W.F. Meyer's Life and Health Insurance Law, A Summary (1976).

  Keith M. Andrews looks at regulation in the United States in Chapter 10, while David B. Martin studies the Canadian situation in Chapter 11. Chapter 12, by Edward P. Potanka, is devoted to regulation of HMOs, PPOs, and managed care in the United States.

- **Section 4—"Underwriting and Managing the Risk,"** W. Duane Kidwell, editor. There is a gap in this section between the large groups of Chapter 13 (as few as 50 or 100 employees) and the small groups of Chapter 15 (under 25 employees). The group underwriter traditionally has made decisions based on qualitative opinions. Given the advances in actuarial modeling, software, computer technology, and data gathering, however, it is time for group insurance underwriting to be based on quantitative data. Neither these chapters nor a later chapter (Chapter 26) devoted to data sources and structure discusses this issue.

  Chapter 13 covers large group underwriting. The authors are James T. Lundberg and Jean C. McFadden.

  The introduction to Chapter 14, "Underwriting Small Groups," by Barbara Niehus, includes some statistics on the extent of insurance in small employers. These statistics appear without citation of source.

  Chapter 15, "Managing Multiple-Choice Situations," by Scott M. Snow, can be supplemented with Fuhrer and Shapiro (1992) and Gifford and Seltz (1988).

  Raymond F. McCaskey covers claim administration and management in Chapter 16.

- **Section 5—"Funding and Rating,"** Francis T. O'Grady, editor. One section of Chapter 17, "Estimating Claim Costs for Life Benefits," by Stephen T. Carter, deals with the effective date adjustment. This adjustment factor is used to adjust for the fact that the manual claim table is set to be correct for rates effective on July 1 based on calendar year of birth ages. If the rates are effective on another date, all employees will be a few months older or younger. The adjustment factor is set at approximately 0.5 percent per month. This, of course, is equal to the weighted average of monthly increases in mortality weighted over the ages of a typical employee group. With the availability of mod-
ern data processing equipment, there is no reason to use this weighted average. Instead, an effective date interpolation can be done for each age. The extra accuracy may not be of the utmost importance, but it costs almost nothing because the basic age/sex rating usually is done by computer.

The book (Chapters 17, 22, or 25) contains a brief treatment of credibility for group life insurance. If the standard assumption is made that all of the experience is equally relevant, then credibility can be shown to equal $ef/(ef+K)$. Here $e$ is the expected number of claims in the experience period, $f$ is an adjustment factor to reduce the credibility for variation in the size of benefits so that $f = 1/(1+\nu/b^2)$, and $\nu$ and $b$ are the variance and expectation of the benefits given a claim has occurred. Note that $f = 1$ if all benefits are the same; otherwise, $f < 1$. This formula assumes, as is usually done, that credibility is applied against total dollars of incurred claims. A better way would be to apply credibility to the number of claims, in which case $f = 1$. Here $K$ is a constant whose value is probably in the 3 to 12 range. This is the constant $k$ in formula (4) on page 525. Herzog explains how $K$ could be estimated on page 530.

Chapter 18, "Estimating Claim Costs for Traditional Health Benefits," by Susan J. Comstock, contains almost no discussion of the experience rating of health benefit claim costs. The method of using claim (charge) experience to build a probability distribution for determining the cost-sharing impact (i.e., deductibles, coinsurance, out-of-pocket maximum, and plan maximum) on pages 333-336 is not optimal. Unfortunately, this method is used by most health insurance actuaries. In this method the charge data are put into size ranges. The probability distribution is defined as a discrete distribution with points set at the average of the charges in each range. The probability is set equal to the number of charges that fall into the range divided by the total number of claims. A simple calculation will show that this method understates the cost for all deductibles except those that fall exactly at a range boundary. For deductibles at a range boundary (or for deductibles that fall in a range with zero or one charge in it) the cost matches the data. Another method is suggested in another context by Gerber and Jones (1976). A better method might be to use the full charge data. The methods of Hogg and Klugman (1984) could be used. Also see Lowrie and Lipsky (1990).

Lee E. Launer details in Chapter 19 various ways to calculate premiums for managed care plans.

Readers interested in reading further about the topics covered in Chapter 20, "Estimating Claim Costs for Disability Benefits," by John C. Antliff and Roy Goldman, should see the discussions of Roy Goldman's paper (1990) for more detail on credibility calculations for LTD.

There is considerable literature on the general business problem of pricing products based on internal expenses and market conditions. There should be some information in Chapter 21, "Calculating Gross Premium and Contribution Rates," by Richard
S. Wolf, on the problem of determining item expenses from expense studies in the field of cost accounting.

Chapter 22, "Experience Rating and Funding Methods," by William F. Bluhm, is similar to his (Bluhm, 1989) study note of the same name. I will discuss a few of the points he raises in this chapter. For example, Bluhm states (pages 410-411) that one of the theoretical considerations entering the choice of credibility levels is the confidence interval chosen by the insurer. Modern least squares credibility (see Chapter 25) does not use confidence intervals, even implicitly. On page 414 Bluhm correctly states that pooling methods are used in order to dampen random statistical fluctuations to make the rates charged as attractive as possible. Pooling methods, however, also are used in prospective experience rating to make the claim projections more accurate. See Fuhrer (1988a) for a method of setting individual claim pooling levels to optimize the calculation of claim cost levels.

For more information on group credibility see Fuhrer (1988a) and compare this to Margolin (1971). For a good method of calculating deficit risk charges, see Panjer and Merleu (1980). Bowers, Gerber, Hickman, Jones, and Nesbitt (1986), Fuhrer (1988b), and Panjer (1980) are good sources for more on calculating aggregate stop loss premiums. There have been some papers on individual stop loss type insurance (casualty) in some of the other actuarial journals. Lowrie and Lipsky (1990) deal with specific stop loss.

- **Section 6—"Economics and Statistics," Jerry E. Lusk, editor.** There has been considerable work done by economists on the problem of estimating trends and analyzing business cycles. Chapter 23, "Medical Claim Cost Trend Analysis and Underwriting Gain/Loss Cycles," by John P. Cookson, continues this body of work.

It would have been useful to include times series extrapolation in Chapter 24, "Forecasting," by Bruce C. MacLeish.

Chapter 25, "Bayesian Statistics and Credibility," by Thomas N. Herzog, is similar to the Transactions of the Society of Actuaries (TSA) paper by the author (Herzog, 1989 with discussion). The TSA discussions of the paper contain valuable info and are quite insightful. The chapter concerns credibility, especially as it relates to Bayesian statistics. There are many books on Bayesian statistics; see, for example, Berger (1985).

- **Section 7—"Information and Its Uses," William F. Bluhm, editor.** Chapter 26 is devoted to the topic "Data Sources and Structures." It was written by Randall P. Herman. Chapter 27, "Management Information Systems," was authored by William R. Lane.

Many books and papers (including Chapter 28, "Claim Reserves," by Mark E. Litow) have been written on insurance claim reserves. I include a bibliography with over 60 entries on this subject in my discussion of the author's paper (Litow, 1989). The development method described by the author has many variations and is only one of myriad methods.
Chapter 29, “Group Insurance Financial Accounting,” by James T. Blackledge, J. Harvey Campbell, and Pierre Saddik, is one of a number of books and articles on insurance accounting. The interested reader should see Saunders (1986).

- **Section 8—“Management,” Bertram N. Pike, editor.** Chapter 30, “Strategic Issues,” was contributed by Donald M. Charsky. Pike discusses the strategic issues facing corporations in general, not the group insurance industry in particular. There are many references to these issues in the general business literature.


  Chapter 32, by Irwin J. Stricker, is devoted to product development.

  Chapter 33, “Organization Structures” by James P. Galasso could be supplemented with references in the general business section of any library related to this subject.

  In Chapter 34, Francis G. Morewood details planning and control issues.

In summary, *Group Insurance* includes many chapters that provide an excellent pedagogy. I hope that future editions will contain a more complete list of references.

**References**


Charles Fuhrer
1630 Chicago Avenue
Evanston IL 60201-3885