1994

*Journal of Actuarial Practice, Volume 2, No. 1, 1994*

Colin Ramsay, Editor
*University of Nebraska - Lincoln, cramsay@unl.edu*

Follow this and additional works at: [http://digitalcommons.unl.edu/joap](http://digitalcommons.unl.edu/joap)

Part of the *Accounting Commons*, *Business Administration, Management, and Operations Commons*, *Corporate Finance Commons*, *Finance and Financial Management Commons*, *Insurance Commons*, and the *Management Sciences and Quantitative Methods Commons*


This Article is brought to you for free and open access by the Finance Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Journal of Actuarial Practice 1993-2006 by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Health Insurance Reform and its Effects on the Small Employer Market: A Review of H.R. 3626
P. Anthony Hammond .......................................................... 5

An Introduction to Individual Disability Income Insurance
Mark J. Chartier ............................................................................................ 47

Cost Containment in Workers’ Compensation:
Evaluating Medical Fee Schedules
David L. Durbin and Barry I. Llewellyn ...................................................... 81

The Markov Chain Interest Rate Scenario Generator Revisited
Sarah L.M. Christiansen ........................................................................... 101

Managing the Costs and Risks of Housing Finance:
A New Role For Actuaries
Anthony Asher .......................................................................................... 125

Reconciling Two Rate Level Indications: A Chain Rule Approach
Cheng-Sheng Peter Wu ............................................................................... 145

Tax Assistance to Qualified Retirement Savings Plans:
Deferral or Waiver?
Robert L. Brown ...................................................................................... 159
EDITORIAL POLICY

The aim of this international journal is to publish articles pertaining to the “art” and/or “science” involved in contemporary actuarial practice.

The Journal welcomes articles providing new ideas, strategies or techniques (or articles improving existing ones) that can be used by practicing actuaries. In addition, the Journal of Actuarial Practice provides a forum for the presentation and discussion of issues of interest to actuaries. One of the goals of the Journal of Actuarial Practice 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
- reinsurance
- reserving issues
- risk-based capital
- risk theory
- social insurance
- solvency issues
- taxation
- valuation issues
- workers’ compensation

REVIEW PROCESS

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 paper is 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 subject to minor revisions, (2) accept contingent on substantive revisions, (3) resubmit: return to the author(s) for major revisions and subsequent resubmission, or (4) 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.
Health Insurance Reform and Its Effects on the Small Employer Market: A Review of H.R. 3626

P. Anthony Hammond*

Abstract†

This paper provides a detailed analysis of H.R. 3626, a bill that is intended to improve employers' and employees' access to health care. H.R. 3626 attempts to accomplish this through the use of guaranteed availability, community rating, and generous standard benefits. A migration model is used to analyze the impact of H.R. 3626. Using this model, it is shown that while improving the availability and affordability of health insurance, its rating restrictions increase premiums disproportionately for the majority of small employers. In addition, H.R. 3626 increases the number of uninsured small employers.

Key words and phrases: rating restrictions, community rating, cost containment, redistributional effects, migration effects

1 Introduction

The majority of Americans obtain their health insurance coverage through an employer-provided health plan. In spite of this fact, many

---

*Anthony Hammond, A.S.A., M.A.A.A., is an associate director and actuary policy at the Health Insurance Association of America. His responsibilities include assessing the costs and implications of federal and state health care reform proposals.

Mr. Hammond's address is: Health Association of America, 1025 Connecticut Avenue NW, Washington DC 20036-3998, USA.

Mr. Hammond thanks the editor and the anonymous referees for their numerous helpful comments and suggestions.

†This paper is a revision of an analysis prepared for the Health Insurance Association of America. Given the current debate on national health care reform, I hope to provide some insights into the methodologies that may be used to analyze the impact of health insurance market reforms and impart an appreciation for the actuarial complexity inherent in these reforms and the consequences of these reforms for insurers, employers, and employees.
working Americans are still uninsured, especially those who work for small employers.\textsuperscript{1} One may argue that the main reason why these workers are uninsured may have more to do with economics than insurance. Yet there are problems in the small employer health insurance market that are exacerbating the problem of the working uninsured. Basically, these are problems of access, affordability, and coverage.

Several small group reform proposals dealing with these problems have been presented to the United States Congress.\textsuperscript{2} These proposals generally promulgate restrictions on health insurance premiums and cost containment measures to improve affordability; require insurers to issue and guarantee renewal of policies to small employer groups in order to improve access (availability) of health insurance; and provide portability (continuity) of coverage when employers or individuals change carriers or jobs or when insurers leave the market.

In spite of their similarities, however, the various proposals are often quite different in their specific provisions. One major difference is how much premiums are allowed to vary (rating restrictions) and the definition of a small employer. The question remains, however, as to how effective these proposals are at resolving some or all of the problems in the small employer market and whether the cost exceeds the benefits to small employers, their employees, insurers, and society as a whole.

This paper examines the efficacy of H.R. 3626, \textit{The Health Insurance Reform and Cost Control Act of 1991}. H.R. 3626 includes all of the approaches mentioned above and goes one step further. It establishes a minimum standard for benefits that must be covered by a small employer health insurance plan. H.R. 3626 does not include, however, reforms such as health risk adjusters, employer mandates, or individual health insurance reforms. Discussion of these reforms is beyond the scope of this paper.

The ability of H.R. 3626 to improve access, affordability, and coverage in the small employer market is analyzed using data from Health Insurance Association of America\textsuperscript{3} (HIAA) member companies and two

---

\textsuperscript{1}A \textit{small employer} is defined throughout this paper to be an employer that employs two to 50 employees.

\textsuperscript{2}As of October 1994, the United States Congress has been unable to pass comprehensive health care reform legislation. So the focus of the health care reform debate has returned to insurance market reforms such as those proposed just a few years ago for small employers. Although these reforms are called incremental by policymakers, they will have a considerable impact on insurers, as these reforms represent a significant departure from past insurance practices.

\textsuperscript{3}The Health Insurance Association of America (HIAA) is a Washington, D.C.-based trade association of the United States' leading commercial insurance carriers. HIAA
actuarial models. One model analyzes the impact of rating restrictions on a sample of small employers insured by five different commercial insurers. The other model (called a migration model) combines the relative morbidity (net claim cost) of each segment of the changing small employer population in order to estimate the change (as a result of H.R. 3626) in the average premium of all small employers insured by small group insurers.

The insurers chosen for this study represent five insurers with significant sales in the commercial, small employer, group health insurance market. The insurers reflect the broad spectrum of underwriting practices that are experienced in the market: from carriers with demographic rating (premiums vary based on age, sex, area, family status) to carriers with aggressive underwriting. Further, while an effort is made to obtain data from a group of carriers that is representative of the small employer group market, it is not possible to determine accurately how representative these carriers are. Therefore, the estimates should not be considered as industry estimates but as the composite experience of five insurers. The results presented in this paper are averages, so it always should be kept in mind that specific insurers and employers will have results that will be higher or lower than the average. In addition, it must be pointed out that small group reforms that already have been implemented in several states will limit the impact of implementing reforms on a national level.

I have tried to make this paper as detailed as possible, but the complexity of small employer market reforms contained in H.R. 3626 and other proposals have exceeded the available data. In response, I have concentrated on those areas that will have the greatest impact on the small employer market, highlighting those factors and effects that will be of greatest concern to policymakers, small employers, insurers, and the small employer population.

Findings from this actuarial study and their implications are presented in detail in the next five sections. Section 2 gives an overview of the results of this paper. Section 3 describes the redistribution of small employer premiums as a result of H.R. 3626 rating restrictions. Section 4 details the changes in the insured and uninsured small employer populations and the effect these population changes will have on small employer premiums. The impact on rates of standardized benefits is described in Section 5. Section 6 covers the impact of cost containment provisions. Section 7 deals with the provisions that can-

represents the majority of the nation's commercial insurance companies. HIAA's activities range broadly from education to legislative analysis to collecting and disseminating data and information.
not be quantified; it also gives suggestions for areas of further study. Section 8 contains the summary and conclusions. Appendix A contains a summary of the basic provisions of H.R. 3626. Appendix B contains a summary of the provisions of H.R. 3626 that affect rating. Appendix C describes the assumptions of the actuarial model under optimistic, best estimate, and pessimistic scenarios; and gives a quantitative evaluation (based on my best estimate) of H.R. 3626's impact on small employers.

2 Overview of the Impact of H.R. 3626

If enacted, H.R. 3626 will make significant changes to the small employer (two to 50 employees) market for health insurance.4 In particular, it will:

- Guarantee that every small employer will have access to coverage;
- Guarantee that all employees (working at least 17.5 hours a week for a small employer with a health insurance plan) and their dependents will be eligible to participate in the employer-provided plan; and
- Make health insurance more affordable for higher risk small employers (thereby providing coverage to more high risk uninsureds).

But it also will

- Make health insurance less affordable for the majority of small employers (more than three-quarters of small employers will receive rate increases of 10 percent or more; see Table 1). For example, Table 1 shows that 19 percent of employees will receive a rate increase of more than 35 percent, 13 percent will retain their coverage, and 6 percent will choose to drop their coverage;
- Increase the small employer average premium per employee 8 to 24 percent, on average, adding an estimated three to nine billion dollars to small employer costs.5 This increase in the average premium is in addition to the rate increases most small employers will receive as a result of rating restrictions. Some small employers, however, will receive decreases in rates; and

---

4This study only addresses the impact of H.R. 3626 on the small employer market, but H.R. 3626 also sets forth portability requirements that apply to all group health plans.

5My best estimate is that premiums will rise 12.3 percent, adding $5.6 billion to small employer costs.
Hammond: H.R. 3626, Health Insurance Reform

Table 1
Distribution of Rate Changes for Currently Insured Small Employers (two to 25 Employees) and the Percentage of Employees Keeping or Dropping Coverage Under H.R. 3626

<table>
<thead>
<tr>
<th>Percentage Rate Change*</th>
<th>Keep</th>
<th>Drop</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 35%</td>
<td>13%</td>
<td>6%</td>
<td>19%</td>
</tr>
<tr>
<td>19% to 35%</td>
<td>23%</td>
<td>3%</td>
<td>26%</td>
</tr>
<tr>
<td>7% to 19%</td>
<td>33%</td>
<td>1%</td>
<td>34%</td>
</tr>
<tr>
<td>-10% to 7%</td>
<td>13%</td>
<td>0%</td>
<td>13%</td>
</tr>
<tr>
<td>Less than -10%</td>
<td>8%</td>
<td>0%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Source: Health Insurance Association of America.

*Includes 12.3 percent (best estimate) increase in market average rate.

- Increase the total number of uninsureds 2 to 5 percent, adding an estimated one to two million persons to the total uninsured. This occurs in spite of the one to two million uninsureds who rejoin the market. These new additions are offset by the one to four million (mostly low risk) employers and employees who leave the market. In addition, the tendency will be for these new uninsureds to be younger, to have lower incomes, and to work for the smallest of the small employers. Many will be children.

The percentage of employers receiving rate increases and the magnitude of those increases are related directly to the degree of rate compression created by rating restrictions. Consequently, the nearly flat community rating of H.R. 3626 leads to more and greater rate increases for employers than might other, less restrictive proposals.

Furthermore, these rate increases are in addition to trend increases and are a direct result of the combination of the access, rating, and benefit provisions of H.R. 3626. (See Table A1 in Appendix A for a summary of these provisions.) H.R. 3626 also will lead to significant changes in who will be insured in the small employer market. Some of these changes are described briefly below.

- **Rating Restrictions:** Under H.R. 3626 rating restrictions, the premium increase experienced by individual small employers will vary widely. Rating restrictions will increase rates significantly.

---

6My best estimate is that the total number of uninsureds will increase 3.4 percent, adding an estimated 1.3 million persons to the total uninsured.
for two-thirds of the currently insured small employers and their employees. Lower income employers and younger employees will be forced to subsidize higher income employers and older employees. Premiums no longer will reflect expected claims, except in the aggregate.

Rating restrictions also lead to a larger percentage rate increase for the smaller small employers than for the larger small employers. This probably reflects two factors: the greater likelihood of the smaller small employers purchasing coverage only if they have a lower than average risk (and, therefore, premium) and the impact of insurer underwriting.

- **Changes in the Insured, Small Employer Population:** The combined results of the H.R. 3626 rating, access, and benefit provisions will be to make health insurance more affordable and accessible for higher risk groups and less affordable for average and lower risk groups. This will lead to adverse selection, i.e., persons who have higher than average health care costs will tend to purchase insurance and those who have lower than average costs will tend not to do so.

Lower risk employers who don't want to drop their coverage also may switch to other forms of coverage that now may be less costly (as a result of H.R. 3626) than group insurance. In addition, the tight rating bands of H.R. 3626 will result in more adverse selection than proposals with less severe rating bands. Thus, H.R. 3626 leads to greater changes in the insured, small employer population than other proposals might. Altogether, H.R. 3626 leads to an 8 to 24 percent average increase in the average premium for small employers and to fewer small employers and their employees being insured.

- **Standardized Benefits:** H.R. 3626 standardizes benefits for small employer plans by preempting state mandates and promulgating a standard benefit package. The standard benefit package is similar to Parts A and B of Medicare, but it also includes certain preventive services with first dollar coverage.\(^7\)

H.R. 3626 increases the self-employment deduction for health insurance and adds four portability provisions that will apply to all group health plans, regardless of size, including self-insured

\(^7\) *First dollar coverage* refers to coverages with no deductible or coinsurance paid by the insured. All charges are fully covered by the insurer.
plans. The portability requirements are: (i) an excise tax (25 percent of gross premium for the plan) for failure to provide all of these portability benefits; (ii) a prohibition against denying, limiting, or conditioning coverage (or benefits) on health status; (iii) a maximum six month preexisting condition limitation (except for newborns); and (iv) a continuity of coverage provision that mandates credit for prior coverage if no more than a three month break in coverage has occurred.

The combination of these benefits is expected to increase premiums about 4 to 5 percent overall for small employers because these benefits, in aggregate, are more generous than the average plan of benefits that small employers currently offer.

- **Cost Containment Provisions:** H.R. 3626 calls for the establishment of a national health care cost containment commission soon after the bill's enactment. It also requires the Secretary of Health and Human Services to develop optional, maximum payment rates for hospitals, physicians, and other health services by October 1, 1994 and annually thereafter. The rates are to be based on DRG (diagnosis-related group) and RBRVS (resource-based relative value scale) methodologies similar to what Medicare currently uses.

These cost containment provisions are too nebulous to justify any estimated reduction in costs at this time. While some studies have estimated significant savings from using current Medicare reimbursement maximums, it is by no means certain that the payment rates eventually approved will be so low. To the extent that the optional DRG and RBRVS rates are used uniformly by health care payors, including government, however, some reductions in cost shifting may occur.

Thus, although H.R. 3626 will improve the availability of coverage for small employers and portability of coverage for all employees, the severe rating restrictions in a voluntary market (without mandated universal coverage) will lead to more persons being uninsured than at present. It will force many small employers to pay a high price to make coverage more affordable for a few small employers. In short, its costs will far exceed its benefits.
3 Rating Restrictions

3.1 The Redistribution of Small Employer Premiums

Most of the provisions in H.R. 3626 that will have a direct impact upon rating are summarized in Appendix B. H.R. 3626 calls for community rating of all small employer (two to 50 employees) health insurance plans with only a ±25 percent adjustment in rates for age and gender if applied consistently to all small employers.

The redistributional effects (or first order effects as they sometimes are called) are the effects of applying the H.R. 3626 rating restrictions to premiums for currently insured employers before any changes occur in the insured population. That is, before anyone migrates to or from (enters or leaves) the small employer market.

Rating restrictions limit the range of premiums that insurers can charge small employers. They redistribute premium rates charged to employers about the average (mean) rate, but the mean rate remains unchanged. As a result, premiums will increase for some employers and decrease for others.

In the discussion of the redistributional effects of H.R. 3626's rating restrictions, the following must be kept in mind:

1. The insured population is held constant when examining the effect of H.R. 3626 rating restrictions on currently insured, small employer groups;

2. The redistributional effects do not include the effects of changes in the insured and uninsured small employer population. The changes in the insured and uninsured small employer population are in addition to the redistributional effects described in this section;

3. The aggregate premium generated from all small employer groups is assumed to be the same before and after rating restrictions are applied. Thus, the average premium is not changed by the effects of H.R. 3626.

4. The change in small employer premium due to redistributional effects is the premium the employer pays after rating restrictions less the premium it paid before rating restrictions. The after premium must not include any increases due to the trend increase employers will receive at renewal or any other increases resulting from H.R. 3626.
5. By definition, there is no redistribuational effect on the uninsured because it is a first order effect, i.e., before migration.

3.2 Methodology

The estimates given in this section are derived from an analysis of a representative sample of actual small employer group data from five different HIAA member companies. Data are collected for similar fee-for-service, indemnity benefit plans (similar to a $200 deductible, 80 percent coinsurance plan), a representative mix of employers for each insurer, and a representative mix by age/sex, industry, area, size, and other small group rating factors for each insurer.

The data are normalized for each insurer before being run through an actuarial model that recalculates the premium each insurer charges each of the 3,750 small employers in the sample using the H.R. 3626 rating restrictions. Geographic factors also are normalized for each insurer, but otherwise are unaffected across insurers. The total premium for each insurer is not changed, but the premium for each employer group is restricted to the H.R. 3626 rating bands such that some employers receive increases and others receive decreases.

The insurers chosen for this study represent five insurers with significant sales in the commercial, small employer, group health insurance market. This group includes insurers with broad and tight underwriting practices. While aggregated estimates are provided, there are large variations between insurers. This suggests that the effect of rate limits will vary greatly from one insurer to another. Further, while an effort is made to obtain data from a group of carriers that will be fairly representative of the small employer group market, there is no way to determine accurately how representative these carriers are. Therefore, the estimates should not be considered industry estimates but should be considered as the composite experience of five insurers.

Representative databases of groups with two to nine employees and ten to 25 employees are obtained from each of the five insurance companies. The two to nine and ten to 25 data are analyzed separately and then combined. The results for groups with two to nine employees are in the same direction but more pronounced than the results for the combined market (i.e., groups with two to 25 employees). Results are somewhat less pronounced for groups with ten to 25 employees.

The database includes employer groups with two to 25 employees rather than groups with two to 50 employees (the definition of a small employer in H.R. 3626). But comparing the effect of H.R. 3626 on groups with two to nine employees versus its effect on groups with ten to 25
employees indicates that H.R. 3626 has relatively less impact on the ten to 25 employee groups. Thus, it seems reasonable to conclude that the effect of H.R. 3626 on groups with 26 to 50 employees will be even less. Hence, groups with two to 25 can serve effectively as proxies for the two to 50 employee groups without significantly affecting the conclusions of this study.

3.3 Impact on Small Employers

Even though the overall average premium remains unchanged, the premium per capita\(^8\) for almost every employer group either will increase or decrease as a result of H.R. 3626. Therefore, the distribution of rates will change. The change in the distribution of premium rates for insured employer groups before rating restrictions, and before migration and expanded benefits, is illustrated in Figures 1 and 2. Figure 1 shows the distribution of premium rates for insured employer groups before rating restrictions, and before migration and expanded benefits (i.e., before H.R. 3626) as a percentage of the mean rate, while Figure 2 shows the distribution after H.R. 3626 rating restrictions are applied.

The impact of the H.R. 3626 rating restrictions on individual small employer groups is examined by determining the premium increase (or decrease) each employer will receive under H.R. 3626’s rating restrictions. The employers are grouped into five categories based on the level of the employer’s percentage change from its current premiums. Then the subsidy provided by low average age employers to high average age employers is examined. Last, the subsidy provided by the smallest small employers to larger small employers and the price sensitivity of the smallest small employers are examined.

Rating restrictions will affect each small employer differently. H.R. 3626 restricts small employer rates to a narrow band, forcing all but a handful of employers to receive premium increases or decreases. These first order effects on the total premium for each small employer in the sample are illustrated in Tables 2, 3, and 4. Table 2, for example, shows that 19 percent of employers with two to 25 employees will have an increase in premium of more than 20 percent from H.R. 3626 rating restrictions alone. For this 19 percent of employers, however, increases will range from 21 percent to 238 percent, and the average increase for all employers receiving more than a 20 percent increase will be 36.6 percent. More than two-thirds of all employer groups (68 percent) will

\(^8\)The premium per capita is the total premium for the employer group divided by the total number of employees and dependents covered by the employer.
receive premium increases from rating restrictions. Nearly half of all employer groups (45 percent) will receive premium increases greater than 5 percent.

Table 2
Distribution of Rate Changes for Currently Insured Small Employers (two to 25 Employees) Under H.R. 3626 Before Migration and Expanded Benefits

<table>
<thead>
<tr>
<th>Percentage Rate Change</th>
<th>% of Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 20%</td>
<td>19%</td>
</tr>
<tr>
<td>6% to 20%</td>
<td>26%</td>
</tr>
<tr>
<td>-6% to 6%</td>
<td>34%</td>
</tr>
<tr>
<td>-20% to -6%</td>
<td>13%</td>
</tr>
<tr>
<td>Less than -20%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Source: Health Insurance Association of America.
To examine the extent to which low average age employers subsidize high average age employers, increases in premiums for the one-fifth (the fifth quintile) of employers with the lowest average employee age are compared to the increases in premiums for all small employers. If the percentage of employers receiving a premium increase is similar in each premium-increase category, it indicates that little or no extra subsidy is demanded from the employers with a younger group of employees. The results, however, show that considerably more low average age employers will receive premium increases than will all employers. Compared to the 68 percent of employers receiving premium increases among all employers, 82 percent of low average age employers are expected to receive premium increases.

Table 3 shows that, in contrast to the 19 percent of employers that will receive a premium increase of more than 20 percent, 30 percent of low average age employers will receive premium increases of this magnitude. Additionally, only 10 percent of low average age employers
will receive a significant premium decrease versus 21 percent of all employers.

Table 3

<table>
<thead>
<tr>
<th>Percentage Rate Change</th>
<th>% of Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 20%</td>
<td>30%</td>
</tr>
<tr>
<td>6% to 20%</td>
<td>30%</td>
</tr>
<tr>
<td>-6% to 6%</td>
<td>30%</td>
</tr>
<tr>
<td>-20% to -6%</td>
<td>7%</td>
</tr>
<tr>
<td>Less than -20%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Source: Health Insurance Association of America.

While the direction of this result is not surprising, its magnitude is significant and indicates the extent to which H.R. 3626 shifts rates from the actuarial goal of consistency between premiums and risk assumed (i.e., premiums no longer reasonably reflect expected claims, except in the aggregate). Also, because of the clear relationship between higher ages and higher incomes, this has the perverse effect of forcing lower income employees to subsidize higher income employees. Thus, the redistributional effect of H.R. 3626 is regressive.

To examine the extent to which the smallest of the small employers will subsidize larger small employers, increases in premiums for the largest size quintile of employers insured by each carrier (generally, employers with more than 15 employees) are compared to the increases in premiums for all small employers. Again, if the percentage of employers receiving a premium increase is similar in each premium increase category, it indicates that little or no extra subsidy exists. The results of this model, however, indicate that considerably fewer of the larger small employers will receive premium increases when compared to all small employers. Compared to the 68 percent of all employers that will receive increases (increases ranging up to 238 percent), only 59 percent of larger small employers will receive increases (increases ranging only as high as 131 percent).

Table 4 shows that 12 percent of these larger small employers (versus 19 percent of all employers) will receive a premium increase of more than 20 percent. Likewise, in contrast to the overall statistic that 32 percent of small employers will receive a premium decrease, 41 percent of
the larger small employers will receive a decrease. That is, larger small employers are almost one-third more likely to receive decreases than the smallest of the small employers.

Table 4

<table>
<thead>
<tr>
<th>Distribution of Rate Changes for the Largest Quintile of Small Employers (two to 25 Employees) Under H.R. 3626 Before Migration and Expanded Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Rate Change</td>
</tr>
<tr>
<td>More than 20%</td>
</tr>
<tr>
<td>6% to 20%</td>
</tr>
<tr>
<td>-6% to 6%</td>
</tr>
<tr>
<td>-20% to -6%</td>
</tr>
<tr>
<td>Less than -20%</td>
</tr>
</tbody>
</table>

Source: Health Insurance Association of America.

There are several ways that Table 4 can be interpreted. The most simplistic way is that the smallest employers pay less because carriers underwrite the smallest employers more aggressively and select better risks, on average. If this is true, the fifth quintile of small employers will have the greatest increases. Upon reviewing the data on the fifth quintile of small employers and comparing it to all small employers, however, the results are ambiguous.

Alternatively, Table 4 can be interpreted as an illustration of the greater price sensitivity of the smallest of the small employers. Because the larger small employers have about the same percentage of premium decreases as all small employers, those larger small employers seem as willing as all small employers to pay the increases when premiums rise. But a smaller percentage of larger small employers receive premium increases than all small employers. As a result, the smallest small employers appear to be either less willing or unable to purchase coverage unless they have a lower than average risk and, as a consequence, receive a lower than average premium. This indicates the greater price sensitivity of the smallest small employers.

It is probably most reasonable to interpret Table 4 as demonstrating the combined result of underwriting decisions by carriers and the price sensitivity of the smallest small employers.
3.4 Impact on Insurers

All insurers are not affected equally under H.R. 3626. The impact of the H.R. 3626 rating restrictions on insurers varies depending on the mix of high and low cost insureds that the insurer underwrites. For example, in contrast to the 19 percent of employers that will receive a premium increase of more than 20 percent, one insurer in the sample will have 37 percent of employers with increases that large.

The rating restrictions in H.R. 3626 break the actuarial link between the premiums insurers charge and the risks they assume. Thus, premiums no longer will reflect the expected claims the insurer will incur for each class of risk. Solvency and actuarial soundness of rates will be affected any time reforms lead to an environment where it is difficult, if not impossible, to charge rates that reasonably reflect the expected costs the insurer will incur for each class of risk. Therefore, the reforms in this bill could place additional financial stress on insurers and could lead to their financial insolvency.

Insurer rate bands will be based on the insurer's own average rate. Some insurers will have lower average rates and some will have higher average rates. Insurers that, coincidentally or because of underwriting before reform, have insured populations with lower than average risk can be expected to have a lower average (or community) rate. Other insurers could insure a population that has a greater proportion of higher than average risks; their community rate is likely to be higher than the community rate of insurers with lower than average risks. Insurers may be able to absorb some losses for a period of time, but continued deterioration will force them out of the small group business. In addition, other lines may subsidize these losses.

Insurers that have the majority of their business in the small group market may not fare as well as insurers with a more diversified mix of business. Insurers that are only in the health business may not have the option of retaining losses until they can leave the market gracefully. Unrecoverable losses could force them into insolvency, further disrupting the market and threatening coverage for their existing policyholders. Furthermore, this will be particularly disruptive to policyholders in managed care networks and to provider-patient relationships.
4 Changes in Insured/Uninsured Small Employers

4.1 Migration Effects on Small Employers Populations

In this section, the migration effects (also called the second order effects) of H.R. 3626's rating restrictions, access provisions,9 and standardized benefits are examined. The migration effects are the effects of employers and employees entering and leaving the small employer group market as a direct result of the change in access rules and changes in premiums.

The rating restrictions (described in Section 3) limit the premium rate that insurers can charge high risk employers, making coverage more affordable for high cost risks. At the same time, access provisions guarantee employers and employees the right to purchase coverage. This combination encourages insured high risk employers and employees to retain their insurance coverage. The access provisions also encourage uninsured high risk employers and employees to purchase insurance coverage. Guaranteed eligibility provisions assure that all employees (working at least 17.5 hours a week for a small employer with a health insurance plan) and their dependents will be eligible to participate in their employer's plan. Guaranteed renewability provisions ensure that once an employer gets coverage, the employer will not lose it.

In addition, H.R. 3626 increases the self-employed deduction for health insurance benefits, and self-insurance10 is limited. The effects of the self-employed deduction and, to a lesser extent, the self-insurance provision lead to changes in the insured small employer population. The expanded benefits also increase premiums for small employers, thus exacerbating population changes.

In the absence of rating restrictions, H.R. 3626's access provisions will lead to some changes in who is insured. These provisions also will lead to either higher rates for the groups affected, to increases in the insurers' average rate for all small employers, or to a combination of both. When combined with the severe rate compression of the nearly flat community rating and the expanded benefits, however, access provisions will increase rates significantly. These increases either (i) add to the increases the majority of small employer groups receive; (ii) make the

9The H.R. 3626 access provisions also are summarized in Appendix B. The most significant access reforms are guaranteed issue, guaranteed eligibility (whole group), and guaranteed renewability.

10H.R. 3626 limits the efficacy of self-insurance, but it does not prohibit it. Consequently, small employers for whom it is still advantageous to self-insure (even with the 25 percent excise tax) may choose to do so.
employers that receive small decreases instead receive rate increases; or
(iii) reduce the rate decreases that employers receiving larger decreases
will receive.

In the following discussion of the migration effects of H.R. 3626's
rating restrictions, the following should be kept in mind:

1. In this section, the insured population is changing. Also, there is
no mandate on the employer to provide, or on the employee to
obtain, coverage, i.e., it is a voluntary market;

2. The effects of migration are examined using sensitivity tests on
a range of assumptions regarding how many insureds enter and
leave the market and the morbidity (net claim costs) of these mi-
grants;

3. Because carriers are insuring a different population, the average
premium of each employer group and the aggregate premium over
all groups combined may change;

4. The difference in premium for all small employers is determined
by comparing the market average of premiums before reforms to
the market average of premiums after reforms. This difference is
in addition to the trend increase employers may receive in their
premiums and to the redistributional effects;

5. Migration effects also have an impact on the number of uninsured
small employers.

4.2 Methodology

The estimates in this section are derived from sensitivity analysis us-
ing the actuarial migration model described in Appendix C. The model
is a reasonable compromise between simplicity and complexity. It iden-
tifies those factors that need to be measured to understand the impli-
cations of small group reforms. In addition, the model shows which
of the factors have the greatest impact on premium changes as a re-
result of reforms. The model splits the small employer health insurance
market into three blocks: (i) employer-provided insurance (including
self-insured employers); (ii) insurance from any other source; and (iii)
the uninsured.

The model addresses the movement of small employers into and out
of the health insurance market, but it does not attempt to simulate the
effect of employers moving between carriers within the market. While
movement within the market may lower a particular employer's premium, it will not change total costs in the market unless employers reduce their coverage at the same time. In the long run, any premium shortfalls from insufficient rates will show up as future trend rate increases.

Starting with the small employer population and the morbidity pattern of each block before reforms, the population and morbidity of insured small employers entering and leaving each of these blocks is combined algebraically in the migration model in order to estimate the small employer population and the morbidity pattern in each block after reforms.

The uninsured are segmented further into high and low risk individuals. Varying assumptions are used for how many individuals enter the insured market from these two segments, as high risk individuals and their employers have a greater incentive to enter the market and have a greater impact on the increase in premiums as a result of H.R. 3626. These assumptions are discussed in greater detail in Sections 4.6 and 4.7.

While population-based data are not consistent with developing exact numerical estimates, reasonable ranges can be developed for the model's assumptions. (See Appendix C.) Consequently, it is necessary to perform sensitivity tests on the results using different sets of assumptions, or scenarios, in order to test the full range of possible values for each variable. For purposes of illustration, these are narrowed to three scenarios that bound the full range of outcomes: low cost, most likely, and high cost. The results vary over a wide range and are sensitive to some of the assumptions.

4.3 Impact on Premium per Capita

Migration effects will increase the premium per capita for all small employer groups covered by an insurer equally. The premium per capita for every employer group will increase above what it would have been due to rating restrictions only. Employers scheduled to receive rate increases will get higher increases. Employers due to receive rate decreases will receive smaller decreases or no decreases. The distribution of premiums per capita for insured employer groups will change to reflect the migration effects as well as redistributional effects.

The distribution of premiums per capita for currently insured employer groups after migration will be similar to the distribution in Figure 2 except the average premium (the 100 percent level) will be higher, i.e., Figure 2 shifts to the right for currently insured employers.
4.4 Impact on Small Employers

H.R. 3626’s access provisions, in conjunction with rating and benefit provisions, will lead to population changes that will cause the average premium per capita to rise for all small employers. This increase could range from 8 to 24 percent of current premiums. This range is broad because the estimated results are sensitive to critical assumptions that are used in the migration model. The upper and lower ends of the range can be considered extremes such that the most likely outcome falls somewhere close to the middle of the range.

The premium increase from migration will be in addition to any increase or decrease that employers will receive from redistributing premiums to meet rating restrictions. For example, if a carrier has to raise the average premium 12 percent for second order effects, any group that previously received a 24 percent increase now will receive a 39 percent increase. (The impact on rates is cumulative, e.g., $1.24 \times 1.12 = 1.3888$, which is roughly a 39 percent increase.) Likewise, any group that had received a 24 percent decrease from rate compression now will receive only a 15 percent decrease ($0.76 \times 1.12 = 0.8512$), effectively wiping out almost half of the benefit of the rating restrictions.

While it is possible to envision better or worse scenarios, it seems most likely that employers will experience premium changes from $-25$ percent to $+271$ percent at renewal when the effects of trend and H.R. 3626 are combined (assuming an 18 percent trend factor). Thus, the covariant effects of community rating in conjunction with guaranteed access and expanded benefits effectively could undermine the goals of greater access, affordability, and coverage in a voluntary market.

Some employers may respond to the rate increases caused by H.R. 3626 by dropping their coverage. Historically, however, employers receiving large trend or experience increases have not responded by dropping their coverage. Instead they try to retain their coverage by seeking less costly alternatives such as reducing plan benefits or increasing employee contributions. Failing that, they may seek coverage through the individual health insurance or self-insurance markets to the extent that these markets provide cheaper alternatives after H.R. 3626 reforms.

Employees experiencing large rate increases also will seek lower cost options. Some employees may choose to reduce their health insurance premium by dropping family coverage in favor of single coverage on the employee only or by increasing their deductibles and copayments.
4.5 Impact on Insurers

The H.R. 3626 rating, access, and benefit provisions will make health insurance more affordable and accessible for high risk groups and less affordable for low risk employers. This will increase adverse selection, i.e., persons who know that they have higher than average health care costs will increase their purchase coverage while those who know that they have lower than average costs will reduce their coverage. The tight rating bands of H.R. 3626 cause more adverse selection than other proposals with less severe rating bands. Thus, H.R. 3626 leads to greater adverse selection and greater changes in the insured, small employer population than other proposals to date.

Because low risk insureds tend to shop for coverage (replace their current coverage for lower cost coverage) more than high risk insureds, insurers with lower average rates can be expected to attract more low cost risks than those with higher average rates. This implies that, for competitive/antislection reasons, some companies will not be able to raise their rates high enough to cover expected claim costs. If their in-force business eventually deteriorates to the point that it contains a significantly disproportionate share of high cost insureds, the insurer will be left with two equally poor choices: reduce rates (and hope that the low cost risks will come) or raise rates and experience further deterioration of their claims experience.

Alternatively, those insurers experiencing enrollment losses as a result of employers dropping coverage (especially low risk ones) could be forced to (i) strengthen and apply participation requirements more strictly; (ii) expand self-insurance products for small employers; or (iii) seek reinsurers/partners in order to spread the risk and maintain market share. The impact on insurers of current migrations of both large and small employers toward self-insured, ERISA-protected plans\(^ {11}\) provides strong empirical evidence of this tendency.

Insurers will be subject to the effects of adverse selection (even from employers that maintain their coverage) if employees who will have to contribute toward higher premiums choose to forego coverage instead. The insurer still may cover the employer, but now fewer employees and their dependents will be in the risk pool. It also can be presumed that employees foregoing coverage will be, as a group, lower risk than those remaining insured.

Insurers that guarantee issue coverage to a disproportionate share of high risk insureds will face an additional risk: their small group busi-

\(^{11}\)ERISA plans are self-insured medical plans established by the United States Congress in accordance with the Employee Retirement Income Security Act of 1974.
ness will have a higher percentage of high cost claimants than their rates anticipated. At the same time, rates only can be set within the allowable rate bands, and rate increases are limited. It may become impossible to offer an actuarially sound, competitive rate that attracts a reasonable mix of high and low risk insureds to assure the integrity of the risk pool. Insurers could find their total premium (from all small groups) insufficient to pay their claims. This increases insurer uncertainty that premiums will be sufficient to pay claims.

Risk margins in current rates are based upon the level of uncertainty (the probability that actual costs will vary from expected costs) that exists in the current market. The increased uncertainty of the market after reforms will encourage insurers to increase their risk margins. The magnitude of the increase will depend on each insurer’s specific situation, and it is unlikely that any a priori estimate of its magnitude will be credible.

4.6 Impact on the Uninsured

Estimating the number of uninsureds attached to the small employer market is hindered by having to determine whether an establishment is a small employer or part of a larger firm. For example, six dry cleaners each with ten employees may be part of the same 60 employee firm or they may be six separate ten employee firms—in both cases they will be six establishments. The question of what to do about dependents when both spouses work and both are uninsured, but one works for a small firm and the other works for a large firm is also problematic. Also, the data are not always split into the employer size categories desired for analysis. In spite of these complications, algorithms have been developed that address these issues. Estimates of the number of small employer uninsured range from 11 to 15 million.

Even in the best scenario, indications are that the number of small group uninsureds will increase rather than decrease under H.R. 3626, contrary to the desired goal of this bill. As many as 0.6 million to 2.3 million Americans may reenter the market under H.R. 3626 reforms. An estimated 1.2 million to 4.1 million more Americans may drop or lose their coverage. The net effect will be an increase in the number of uninsureds of 0.7 million to 1.9 million, increasing the number of small group uninsureds 6 to 12 percent. Because small group uninsureds are about half of the uninsured, however, H.R. 3626 will increase the total number of uninsureds about 3 to 6 percent. If previous socioeconomic patterns hold for these new uninsureds, the tendency will be for these new uninsureds to be younger, lower income, and from the smallest
4.7 Number of High Risk Uninsureds

A key assumption is the number of high risk uninsured employees. High risk uninsureds are medically uninsurable individuals who may be denied health insurance under current medical underwriting practices. Prior to a recent AHCPR (Agency for Health Care Policy and Research, United States Department of Health and Human Services) study, reasonable estimates for this variable ranged from 7 to 12 percent. The AHCPR study shows that only 36.8 percent of the uninsured have investigated the cost of private health insurance; only 2.5 percent of that cohort have ever been denied coverage or had their coverage limited. The 2.5 percent includes more than just medically uninsurable individuals and includes those who have been excluded from individual (not just group) coverage. It also includes those who ever have been rejected for a policy, whether they will be today or not. It does present an upper bound for who may be medically uninsurable among the 36.8 percent who have investigated coverage. Assuming the same proportion of uninsurable persons among those who haven't investigated coverage as among those who have (a grossly conservative assumption), 6.8 percent of the uninsured at most could be medically uninsurable.

Medically uninsurable individuals may not be distributed uniformly among the various segments of the uninsured population. For example, small employers could have a higher percentage of medically uninsurable employees than large employers. In case there is a disproportionate share of these high cost insureds among small group uninsureds, 7 percent is assumed to be the low end of the range for small group uninsureds. Even with the possibility of a biased distribution, however, it appears that the 12 percent estimate for the top of the range for this assumption is too conservative. But the 7 to 12 percent range used in this study encompasses the most reasonable range of values available from current research.

4.8 The Morbidity of Medically Uninsurable Employees

Current studies show that the morbidity of medically uninsurable employees range anywhere from 200 percent of the net claim cost of
the current small employer group market to as much as 500 percent. Analysis of the experience of individual high risk pools, however, shows that even when considered as individuals rather than as groups, the experience for these high cost insureds only averages 350 percent of the current small employer group market experience. Because groups can be expected to have employees and dependents that are standard or better risks to offset the additional claim cost of high cost insureds, it is expected that, on average, the morbidity of high risk groups will be less than the morbidity of individual high risk pools. For this study a range of 248 to 300 percent of current small employer net claim cost is used for this variable.

5 Standardized Benefits and Deductions

H.R. 3626 mandates a generous preexisting condition limitation, the preemption of state mandates, the addition of preventive services with first dollar coverage, and a standard benefit package similar to Parts A and B of Medicare. In an attempt to place the self-employed on parity with all other employers, H.R. 3626 increases the self-employed deduction for health care expenses to 100 percent of expenses. These provisions are described in Appendix B.

The standardized benefits detailed below are expected to increase small employer premiums about 4 to 5 percent overall because these benefits are more generous in aggregate than the average plan of benefits small groups currently offer. The impact of the deductions is harder to quantify. The following is a description of the standardized benefits and deductions.

- **Preexisting Condition Limitation:** Based on data from HIAA’s employer survey and HIAA calculations using the 1994 Tillinghast Group Medical Insurance Rate Manual,\(^\text{12}\) it is estimated that reducing the preexisting condition limitation period to a required maximum of six months will add about 2 percent to an average policy.

- **Standard Benefit Package, Including Preventative Services:** The standard benefits package (except for some of the preventive benefits) will be less than a standard employer provided plan in some

\(^{12}\)The *1994 Tillinghast Group Medical Insurance Rate Manual* is published by Tillinghast, 101 South Hanley Street, St. Louis MO 63105-3411, USA.
states, while in other states the H.R. 3626 plan will be more generous. The net effect is estimated to be about a 1 percent increase in the average premium from current levels.

- **Elimination of State Mandates:** The effect of eliminating state mandated benefits and policy provisions is included in the pricing of the standard benefit package. The reduction for eliminating state mandates is not readily apparent because H.R. 3626 mandates a package of benefits that, with the exceptions above, is similar to the average small employer plan. Furthermore, for a small employer that offers its employees a plan with fewer benefits than the H.R. 3626 minimums, rate increases will be even higher than this analysis otherwise indicates.

- **Self-Employed Deduction:** It is difficult to estimate the impact of the self-employed deduction provision. Though the self-employed population is small compared to the total population, it is reasonable to assume some increase. For example, increasing this deduction will tend to encourage the self-employed with above average costs to seek insurance more than it may encourage the self-employed with below average costs to seek insurance. This is evident in the results of the 1987 national medical expenditure survey\(^\text{13}\) that show individually insured persons have much higher cost and risk than group insureds. I estimate that this may add another 1 to 2 percent to the average premium.

### 6 Cost Containment Provisions

H.R. 3626 calls for the establishment of a national health care cost containment commission shortly after enactment of the bill. It also requires the Secretary of the Department of Health and Human Services to establish optional maximum payment rates for hospitals, physicians, and other health services by October 1, 1994 and annually thereafter. The rates are to be based on DRG and RBRVS methodologies similar to those Medicare currently uses.

This approach to cost containment will not contain health care costs effectively because it does nothing to control the fundamental sources of health care cost increases other than medical price inflation.

---

\(^{13}\) The national medical expenditure survey is a detailed survey of the health expenditures of Americans and their families. This survey is sponsored by the US Agency for Health Care Policy and Research, Department of Health and Human Services, 200 Independence Avenue SW, Washington DC 20201.
The most important sources of health care cost increases are medical cost inflation, cost shifting, utilization (the number of health care services used), adverse selection, defensive medicine, and new technologies. In 1991 the medical cost inflation component was responsible for only about a third of health care cost increases. Thus, an approach that addresses only medical cost inflation addresses only about one-third of the problem.

But H.R. 3626 may not control medical cost inflation effectively. While some studies have estimated significant savings from using current Medicare reimbursement maximums, it is by no means certain that the payment rates eventually approved will be so low. Because the Secretary of Health and Human Services will be charged with establishing maximum payment rates (without guidelines for how high these rates could be), it is not certain that the maximum payment rates set by the Secretary will contain costs effectively.

The only positive point that can be made regarding this approach is that, to the extent that the maximum payment rates are used uniformly by health care payors (including government payors), some reductions in cost shifting may occur. It is reasonable to expect that insurers will use these rates if they are legislated or if they are less than what the insurer currently pays. Any insurer that does not likely will be at a competitive disadvantage.

The H.R. 3626 cost containment provisions are still too nebulous to justify any estimated reduction in costs at this time.

7 Provisions Not Quantified

The number and complexity of health insurance market reform proposals have outstripped the available data, and H.R. 3626 is no exception. Consequently, it is impossible to quantify certain provisions in the bill. In some cases more data and analysis are needed. In others data are not available to estimate credibly the impact of certain reforms on the market.

Some of the provisions not specifically quantified in this study are the minimum plan period, the notice of renewal, the index rate variation between blocks, the types of family enrollment, the transfers among blocks, the 5 percent limit on rate increases above trend, the geographic factors limited to MSAs, the self-insurance prohibition, the uniform claims forms, and the uniform reporting standards.

14As defined by the United States Census Bureau, MSA means a metropolitan statistical area, e.g., Hartford, Connecticut.
Although the effect of H.R. 3626 on employees purchasing single coverage versus employees purchasing family coverage is not analyzed in this study, an independent study by an HIAA member company shows H.R. 3626 can be expected to increase rates for single coverage more than for family or single parent coverages. (Increases for singles are estimated to be five times greater.)

Some covariant effects cannot be analyzed with the data available and are beyond the scope of this study. For example, how geographic factors may change in the absence of other risk classification factors (such as industry or full age/gender rating) is not examined. As the scope of this study is limited to the effects of H.R. 3626 on the small employer group health insurance market, the impact of portability requirements on employers other than small employers is not analyzed. Similarly, the impact of H.R. 3626 on association groups and employer-provided individual health insurance is also beyond the scope of this study.

No specific attempt is made in this study to measure H.R. 3626's effect on the solvency of employers and insurers. But the magnitude of rate increases for some employers and the likelihood that certain insurers will get a disproportionate share of high risk insureds will have an impact on their solvency. Also, it is not possible to include the impact of state regulations already promulgated. For example, some states have passed laws similar to the rating restrictions and other provisions in this bill. In these states, to the extent that premiums and the market already reflect these changes, H.R. 3626 will have less impact.

8 Summary and Conclusions

H.R. 3626 tries to marry the social goals of guaranteed availability, community rating, and generous standard benefits. It doesn't consider sufficiently the realities of price-sensitive small employers and individual employees acting in their own best interest in a highly competitive, voluntary market.

While improving availability and affordability of health insurance for a minority of small employers, H.R. 3626's rating restrictions increase premiums disproportionately for the majority of small employers. Employers with younger, lower income employees will be forced to subsidize employers with older, higher income employees. Smaller small employers will subsidize larger small employers. Premiums no longer will reflect expected claims, except in the aggregate, exacerbating the tendency in small employer markets to be uninsured due to cost.
Changes in the insured small employer population as a result of H.R. 3626 will increase the average premium per capita of all small employer groups 8 to 24 percent, largely as a result of the severe rating restrictions. The premium per capita for almost every small employer group either will increase or decrease, and the distribution of premiums per capita for insured employer groups will change, generally worsening. In effect, H.R. 3626 proposes a solution that affects 100 percent of insured small employers, most of them negatively, in order to address a problem that afflicts less than 15 percent of small employers.

Some employers will respond to the H.R. 3626 rate increases by dropping their coverage altogether. Based on my experience with selling coverage to small employers, it seems far more likely that small employers who currently have coverage will do what they have always done when faced with significant rate increases. They will seek less costly alternatives such as reducing benefits to the minimum allowable (if their benefits are currently more generous), increasing employee contributions (employee share of the premium), self-insuring (if feasible), or utilizing the individual health insurance market.

While the full impact is not yet clear, the impact of the community rating law for small employers in New York State seems consistent with this conclusion. Young and healthy lives have dropped out of the system, claims costs have risen, and, at least anecdotally, small employers and individuals are choosing less generous benefit plans. (New York does not mandate a minimum or standard benefit plan.) Mitigating any negative impact of reforms in New York State is the implementation of a risk adjustment mechanism for risk pooling across the individual and small group markets. If H.R. 3626 had contained such a provision, the impact on rates and the market would be less.

Employees experiencing large rate increases also will seek lower cost alternatives. Employees could choose to drop their coverage or reduce their contribution toward premiums by dropping family coverage in favor of single coverage on the employee only.

H.R. 3626 will increase the risk to insurers of providing small employer coverage. Coupled with a hostile regulatory environment wherein rate increases often are reduced or denied, insurers will find it increasingly difficult to charge premiums that are adequate to protect existing policyholders. In response, insurers staying in the market may try to strengthen and apply participation requirements more strictly, increase risk margins, expand self-insurance products to small employers, reduce rates below an actuarially sound level (in an attempt to achieve or maintain a standard mix of risks), or develop more innovative responses that will protect the insurer from insolvency.
In a highly competitive, voluntary insurance market, the actuarial process of rating, the underwriting process of risk selection, and the competitive market in which they operate are symbiotic. Actuarially sound rates are established to ensure the solvency of the insurer in order to protect its policyholders. If these rates are too high, some persons will not purchase insurance. If rates are too low, there will not be sufficient reserves to pay policyholder claims. If the insurer screens too many risks, there will not be sufficient policyholders to cover the costs of operating the business. If the insurer screens too few or does not assign the appropriate rate to high risk policyholders, it will not have sufficient reserves to pay policyholder claims. There is always a decision to be made regarding costs versus benefits whenever considering changes to any of these processes—changing one affects all.

H.R. 3626 fails this litmus test of cost/benefit analysis. It does so, in large part, because of its severe rating restrictions. In the current market where health care costs are such a large proportion of nonsalary employee expenses, employers are looking to reduce this expense and are unwilling to subsidize actuarially higher risk insureds of another employer. If their employees are actuarially low risk, they demand low premiums. Otherwise, they do not purchase insurance. Consequently, forcing insurers in a free market to charge premiums that do not reflect the expected claims of insureds (thereby forcing large rate increases on most small employers) in order to satisfy a social goal will not produce the intended result.

This study shows that it is not the guaranteed issue/availability provisions of H.R. 3626 that lead to most of the small employer premium increases; rather it is the bill's rating restrictions. Other small group reform proposals with less severe rating restrictions will provide the benefits of guaranteed availability without the onerous rate increases precipitated by H.R. 3626.

In closing, although H.R. 3626 will improve availability of coverage for small employers and portability of coverage for all employees, the severe rating restrictions will lead to more persons being uninsured. It will force many small employers to pay a high price to make coverage more affordable for a few small employers. In short, the costs of this bill will far exceed its benefits.

<table>
<thead>
<tr>
<th>Provision</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Size</td>
<td>Two to 50 employees (portability provisions apply to all group health plans);</td>
</tr>
<tr>
<td>Transitional Period</td>
<td>Various, but up to three years for some provisions;</td>
</tr>
<tr>
<td>Availability</td>
<td>Guaranteed issue (year round; uniform waiting periods and minimum participation requirements allowed);</td>
</tr>
<tr>
<td>Individual Policies</td>
<td>Not applicable to individual policies (unless provided by employer);</td>
</tr>
<tr>
<td>Case Characteristics</td>
<td>Age, gender, and geography (no smaller than MSA);</td>
</tr>
<tr>
<td>Rating Restrictions</td>
<td>Community rating such that variations between blocks of business shall not exceed 20 percent. Age and sex adjustments may be used, but only up to ±25 percent and only if applied to all small employers;</td>
</tr>
<tr>
<td>Renewal Rating</td>
<td>May not exceed the sum of the percentage change in the base premium rate plus 5 percentage points;</td>
</tr>
<tr>
<td>Renewability</td>
<td>• Guaranteed renewable except for nonpayment of premiums, fraud or misrepresentation, and failure to maintain minimum participation rates;</td>
</tr>
<tr>
<td></td>
<td>• Must give notice 60 days prior to renewal date; terms of renewal must be same as at issue except for premiums and administrative changes;</td>
</tr>
<tr>
<td>Whole Groups</td>
<td>Coverage must be offered to any eligible employee and dependent;</td>
</tr>
<tr>
<td>Reinsurance</td>
<td>Not included;</td>
</tr>
<tr>
<td>Reinsurance Price</td>
<td>Not applicable;</td>
</tr>
<tr>
<td>Cost Sharing</td>
<td>Not applicable;</td>
</tr>
<tr>
<td>Assessments</td>
<td>Not applicable;</td>
</tr>
</tbody>
</table>

*MSA = Metropolitan statistical area.
Table A1 (continued)
Summary of the Basic Benefit Provisions of H.R. 3626

<table>
<thead>
<tr>
<th>Provision</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portability</td>
<td>These provisions apply to all group health plans:</td>
</tr>
<tr>
<td></td>
<td>• Excise tax for failure to provide all of these portability benefits (25 percent of gross premiums);</td>
</tr>
<tr>
<td></td>
<td>• Prohibition against denying, limiting, or conditioning coverage (or benefits) on health status;</td>
</tr>
<tr>
<td></td>
<td>• Maximum six month preexisting condition limitation (except for newborns);</td>
</tr>
<tr>
<td></td>
<td>• Continuity of coverage provision that mandates credit for prior coverage if no more than a three month break in coverage;</td>
</tr>
<tr>
<td>Other</td>
<td>• Self-employed deduction increased to 100 percent;</td>
</tr>
<tr>
<td></td>
<td>• Applies to employees working at least 17.5 hours per week;</td>
</tr>
<tr>
<td></td>
<td>• Deductible standard benefit package with preventive benefits;</td>
</tr>
<tr>
<td></td>
<td>• Preemption of state mandates beyond standard benefit package;</td>
</tr>
<tr>
<td></td>
<td>• Percent excise tax on self-insured;</td>
</tr>
<tr>
<td></td>
<td>• Any payor may choose to use DRG and RBRVS schedules;</td>
</tr>
<tr>
<td></td>
<td>• Must offer single, couple, single parent, and family rates</td>
</tr>
<tr>
<td>Effective Date</td>
<td>Various: depends on provision (some on January 1, 1992).</td>
</tr>
</tbody>
</table>

*DRG = Diagnosis-related group.

*RBRVS = Resource-based relative value scale.
Appendix B: H.R. 3626, Summary of Provisions Affecting Rating

H.R. 3626 has provisions that apply to small groups (defined as employer groups of two to 50 employees) and all employer groups, including self-insured plans. An employee is defined as any worker who normally works at least 17.5 hours per week.

**Self-employed deduction:** The amount of deduction for self-employed individuals is extended indefinitely, starting in 1993. The deduction increases from 25 percent to 50 percent in 1993, 75 percent in 1994, and 100 percent in 1995 and subsequent years.

**Preemption of state-mandated benefits:** States cannot mandate benefits beyond those in the standard benefit package, but they can establish more stringent requirements in other areas.

**Guaranteed eligibility:** Insurer cannot exclude any eligible employee or dependent to whom the employer offers coverage.

**Guaranteed issue:** Insurers offering a plan to small employers in a community must offer it to all employers in the area year round. Waiting periods are allowed if applied to all employees.

**Minimum plan period:** Rating basis applies for 12 months.

**Guaranteed renewability:** Insurers only can nonrenew and cancel for nonpayment of premiums, fraud, misrepresentation, or failure to maintain minimum participation rates.

**Notice of renewal (expiration):** Insurers must give notice 60 days prior to the renewal date. Terms of renewal must be the same as at issue except for premiums and administrative changes.

**Discrimination based on health status:** Insurers cannot deny, limit, or condition coverage or benefits based on an individual's "health status, claims experience, receipt of medical care, medical history or lack of evidence of insurability." An exception is made to this provision to allow for the preexisting condition exclusion.

**Index rate variation between blocks:** This must be less than 20 percent unless the block is one that always has provided open enrollment, the insurer never has transferred groups into the block involuntarily, and the block is currently available for purchase when an exception to the 20 percent rule is sought.
Community rating: Must be used within a block.

Age and sex adjustments: This may be used if applied consistently to all small employers. Maximum variation is ±25 percent.

Definition of community: Not smaller than an MSA.

Types of family enrollment: Insurers must have separate rates for single adults, childless couples, single parents, and families.

Transfers among blocks: Insurer cannot force an employer to transfer among blocks and may not transfer an employer unless the transfer is offered to all small employer plans and unless it is not based on demographics, experience, or date of issue.

Limits on rate increases: Increase may not exceed the percentage change in the base premium rate plus 5 percent (500 basis points).

Definitions:

1. A block (of business) consists of the small employer plans issued by an insurer. Distinct groups can be treated as separate blocks based on whether the group is marketed through direct response, has been acquired from another insurer, or is provided via an association of at least 25 small employers.

2. The reference premium rate is the lowest rate charged or available to any actuarial class.

3. The index rate is 133 1/3 percent of the reference premium rate.

4. The base premium rate, though not specifically defined, can be defined to be the index rate.

Standard benefit package:

1. In general, same as Parts A and B of Medicare.

2. Unlimited inpatient hospital coverage for children without coinsurance. (The deductible is not excluded.)

3. Maternity (including prenatal, inpatient labor and delivery, postnatal, and postnatal family planning).

4. The $250/500 deductible is indexed for future inflation.

5. The individual out-of-pocket limit of $2500/3000 is indexed for inflation in future years.
6. Preventive services must be provided without deductible or coinsurance. These services include: maternity, well-child care (including dental), screening mammography, screening pap smear, colorectal screening, and certain immunizations. Others may be added at a later date. Their are limitations on what providers may charge for these. Effective 1/1/92.

Self-insurance prohibition: Small employers (two to 50 employees, including self-employed) may not self insure. This is enforced through a 25 percent excise tax on health care expenditures by self-insured plans.

Preexisting condition limitation (PCL) for all groups: Limits pre-existing condition exclusion to six months with a further proviso that prior coverage must be credited toward the six months as long as there isn’t more than a three month lapse in coverage. PCL cannot be applied to newborns and is defined as a condition diagnosed or treated during the three months prior to issue. This applies to all employers. Effective 1/1/93.

Other portability provisions: In addition to the PCL, the portability provisions are an excise tax for failure to provide all of these portability benefits (25 percent of gross premium for plan); a prohibition against denying, limiting, or conditioning coverage (or benefits) on health status; and a continuity of coverage provision that mandates credit for prior coverage if no more than a three month break in coverage has occurred. All of the portability provisions apply to all group health plans, regardless of size, including self-insured plans.

Cost containment: This includes optional rates (prices) for hospitals, physicians and other medical providers, DRGs, and RBRVS. Any health care purchaser, including individuals, can choose to use promulgated rates. Providers must accept these rates as payment in full.

Uniform claims forms: Effective 1/1/94.

Uniform reporting standards: For development of rates (prices) to be used in cost containment efforts. Effective 1/1/93.
Appendix C: H.R. 3626: The Impact of Migration

Table C1
Assumptions for Migration Modeling Under the Optimistic, Best Estimate, and Pessimistic Scenarios

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Optimistic</th>
<th>Best Estimate</th>
<th>Pessimistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Market</td>
<td>45,000,000</td>
<td>50,000,000</td>
<td>55,000,000</td>
</tr>
</tbody>
</table>

Distribution (before reforms):
- A. Employer-sponsored 55% 51.36% 50.0%
- B. Individually-insured 20% 22.31% 22.5%
- C. Uninsured 25% 26.33% 27.5%

Morbidity of Population as a Percentage of Employer-Sponsored Net Claims Cost (before migration):
- A. Employer-sponsored 100.00% 100.00% 100%
- B. Individually-insured 100.00% 100.00% 120%
- C. Uninsured 75.04% 80.62% 102%

Employer-Sponsored Insureds Withdrawing From Small Employer Market:
- A. % withdrawing 5% 10% 15%
- B. Morbidity 24% 32% 40%

Individually Insureds Withdrawing From Small Employer Market:
- A. % withdrawing 5% 5% 5%
- B. Morbidity 100% 120% 120%
Table C1 (continued)
Assumptions for Migration Modeling Under the Optimistic, Best Estimate and Pessimistic Scenarios

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Optimistic</th>
<th>Best Estimate</th>
<th>Pessimistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additions of Uninsureds to Small Employer Market:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Percentage of uninsureds who are medically uninsurable</td>
<td>7%</td>
<td>10%</td>
<td>12%</td>
</tr>
<tr>
<td>B. Percentage of medically uninsurables purchasing coverage after reforms</td>
<td>25%</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>C. Percentage of all uninsureds purchasing coverage after reforms</td>
<td>5%</td>
<td>10%</td>
<td>15%</td>
</tr>
<tr>
<td>D. Morbidity of medically uninsurables</td>
<td>248.08%</td>
<td>248.08%</td>
<td>300%</td>
</tr>
<tr>
<td>E. Morbidity of uninsureds who are not medically uninsurable</td>
<td>62.02%</td>
<td>62.02%</td>
<td>75.00%</td>
</tr>
</tbody>
</table>

*Note: Total Market = Population (employees and dependents) in small employer market before any changes*
Table C2
Basic Set of Assumptions

<table>
<thead>
<tr>
<th>Total Market (before migration):</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ER-sponsored Insureds</td>
<td>25,680,000</td>
<td>51.36%</td>
<td>100.00%</td>
<td>$1,470.74</td>
<td>1.95</td>
</tr>
<tr>
<td>B. Individually insured</td>
<td>11,155,000</td>
<td>22.31%</td>
<td>100.00%</td>
<td>$1,470.74</td>
<td>1.95</td>
</tr>
<tr>
<td>C. Uninsured</td>
<td>13,165,000</td>
<td>26.33%</td>
<td>80.62%</td>
<td>$1,185.71</td>
<td>1.45</td>
</tr>
<tr>
<td>TOTAL</td>
<td>50,000,000</td>
<td>100.00%</td>
<td>94.90%</td>
<td>$1,395.69</td>
<td>1.79</td>
</tr>
</tbody>
</table>

Impact of Migration on:

<table>
<thead>
<tr>
<th>ER-Sponsored Insureds:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Withdraw from market</td>
<td>2,568,000</td>
<td>10.0%</td>
<td>32.00%</td>
<td>$470.64</td>
<td>1.95</td>
</tr>
<tr>
<td>B. Remain in market</td>
<td>23,112,000</td>
<td>90.0%</td>
<td>107.56%</td>
<td>$1,581.86</td>
<td>1.95</td>
</tr>
<tr>
<td>Subtotal</td>
<td>25,680,000</td>
<td>100.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individually Insureds:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Migrate to SGM</td>
<td>557,750</td>
<td>5.0%</td>
<td>120.00%</td>
<td>$1,764.89</td>
<td>1.95</td>
</tr>
<tr>
<td>B. Remain individually insured</td>
<td>10,597,250</td>
<td>95.0%</td>
<td>98.95%</td>
<td>$1,455.26</td>
<td>1.95</td>
</tr>
<tr>
<td>Subtotal</td>
<td>11,155,000</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Uninsureds (UIs):

A. MUIs migrating to SGM
   - Insurables entering SGM
     - Insurables entering SGM
   B. Uninsureds who remain

Subtotal

<table>
<thead>
<tr>
<th>Subtotal</th>
<th>Total Market (after migration):</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ER-sponsored insureds</td>
<td>658,250</td>
</tr>
<tr>
<td>B. Individually insured</td>
<td>658,250</td>
</tr>
<tr>
<td>C. Uninsured</td>
<td>11,848,500</td>
</tr>
</tbody>
</table>

Subtotal |

| Subtotal | 13,165,000 |

Note: Column headings are as follows: (1) Number of Covered Lives; (2) Percent of subtotal or total in column (1); (3) Ratio of Net claim Costs to Market Cost (Market Cost = $1,470.74); (4) Cost Per Covered Life; and (5) Average Family Size. Data in columns (2) to (5) are rounded to two decimal places. ER = Employer.

† These withdrawing employer-sponsored insureds are now considered as uninsured.

‡ SGM = Small group market.

§ MUI = Medically uninsured.
Table C3
Calculations Using Data From Table C2

<table>
<thead>
<tr>
<th></th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Market (before migration):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Employer-sponsored insureds</td>
<td>37,768,603,200</td>
<td>13,169,231</td>
<td>$2,867.94</td>
<td>$239.00</td>
</tr>
<tr>
<td>B. Individually insured</td>
<td>16,406,104,700</td>
<td>5,720,513</td>
<td>$2,867.94</td>
<td>$239.00</td>
</tr>
<tr>
<td>C. Uninsured</td>
<td>15,609,879,891</td>
<td>9,079,310</td>
<td>$1,719.28</td>
<td>$143.27</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>69,784,587,791</td>
<td>27,969,054</td>
<td>$2,495.06</td>
<td>$207.92</td>
</tr>
</tbody>
</table>

**Impact of Migration on:**

**Employer-Sponsored Insureds:**
A. Withdraw from market | 1,208,595,302 | 1,316,923 | $917.74 | $76.48
B. Remain in market | 36,560,007,898 | 11,852,308 | $3,084.63 | $257.05

**Individually Insured:**
A. Migrate to small group market | 984,366,282 | 286,026  | $3,441.53 | $286.79
B. Remain individually insured | 15,421,738,418 | 5,434,487 | $2,837.75 | $236.48

**Uninsured:**
A. Uninsured MUIs' migrating | 2,401,698,712 | 453,966  | $5,290.49 | $440.87
   Insurable migrants | 600,424,678  | 453,966  | $1,322.62 | $110.21
   Subtotal migrating | 3,002,123,390 | 907,931  | $3,306.55 | $275.55
B. Uninsureds who remain | 12,607,756,501 | 8,171,379 | $1,542.92 | $128.57
<table>
<thead>
<tr>
<th></th>
<th>Number of Covered Employees</th>
<th>Annual Cost Per Employee</th>
<th>Monthly Cost Per Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Employer-sponsored insureds</td>
<td>40,546,497,570</td>
<td>$3,107.90</td>
<td>$258.99</td>
</tr>
<tr>
<td>B. Individually insured</td>
<td>15,421,738,418</td>
<td>$2,837.75</td>
<td>$236.48</td>
</tr>
<tr>
<td>C. Uninsured</td>
<td>13,816,351,803</td>
<td>$1,456.15</td>
<td>$121.35</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>69,784,587,791</td>
<td>$2,495.06</td>
<td>$207.92</td>
</tr>
</tbody>
</table>

Note: Column headings are as follows: (6) Total Cost; (7) Number of Covered Employees; (8) Annual Cost Per Employee; (9) Monthly Cost Per Employee.

'MUls = Medically uninsureds.

Data in columns (8) and (9) are rounded to two decimal places.
Table C4
Best Estimate of Financial Impact of H.R. 3626 Small Group Reforms: Increase in Employer-Sponsored Insured Small Group Net Claim Cost due to Guarantees and Other Benefits After Reforms (Calculations Using Data From Tables C2 and C3)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guarantees¹</td>
<td>10.3%</td>
<td>$152.01</td>
<td>$2,777,894,370</td>
<td>$239.96</td>
<td>$20.00</td>
</tr>
<tr>
<td>Other Benefits²</td>
<td>4.0%</td>
<td>$58.83</td>
<td>$1,469,931,093</td>
<td>$112.67</td>
<td>$9.39</td>
</tr>
<tr>
<td>TOTAL</td>
<td>14.3%</td>
<td>$210.84</td>
<td>$4,247,825,463</td>
<td>$352.63</td>
<td>$29.39</td>
</tr>
</tbody>
</table>

Percentage Change | 14.3% | 11.2% | 12.3% | 12.3% |

Note: Column headings are as follows: (1) Increase in the Ratio of Net Claim Costs to Market Cost; (2) Increase in the Cost Per Covered Life; (3) Increase in the Total Cost; (4) Increase in the Annual Cost Per Employee; and (5) Increase in the Monthly Cost Per Employee.

¹These include guaranteed issue, eligibility, and renewability; community rating; and rating restrictions.

²These include preexisting condition limits, self-employment deduction, preventive services, elimination of mandates, cost containment, and the standard benefit package.

Column (1): 10.3% is taken from 110.34% in Table C2, column; and 4.0% is based on information in Section 5.

Column (2) = Column (1) x 1470.74.

Column (3): $2,777,894,370 = $40,546,497,570 - $37,768,603,200; and $1,469,931,093 = $58.83 x 24,986,250.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>Total ER</em>-Sponsored (Before Reforms)</em>*</td>
<td>25,680,000</td>
<td>100.00%</td>
<td>$1,470.74</td>
<td>$37,768,603,200</td>
<td>1.95</td>
<td>13,169,231</td>
<td>$2,867.94</td>
<td></td>
</tr>
<tr>
<td><strong>Increases due to:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current insureds</td>
<td>-2,568,000</td>
<td>14.34%</td>
<td>$1,096,120,369</td>
<td>-1,316,924</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrations</td>
<td>1,874,250</td>
<td>14.34%</td>
<td>$3,151,705,095</td>
<td>1,193,957</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>-693,750</td>
<td>14.34%</td>
<td>$4,247,825,463</td>
<td>13,046,264</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total ER-Sponsored (After Reforms)</strong></td>
<td>24,986,250</td>
<td>97.30%</td>
<td>$1,681.58</td>
<td>$42,016,428,663</td>
<td>1.92</td>
<td>13,046,264</td>
<td>$3,220.57</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Column headings are as follows: (1) Number of Covered Lives; (2) Percentage of Total Number of Employer-Sponsored Insureds; (3) Ratio of Net Claim Costs to Market Cost (Market Cost = $1470.74); (4) Cost Per Covered Life; (5) Total Cost; (6) Average Family Size; (7) Number of Covered Employees; and (8) Annual Cost Per Employee.

*ER = Employer.
An Introduction to Individual Disability Income Insurance

Mark J. Chartier*

Abstract

There are several actuarial software packages purporting to calculate expected benefit cash flows on disability income insurance policies. To the author's knowledge, however, there is no published text that explains how to perform these calculations. This paper is intended to fill this gap in the literature. It describes some of the more common techniques for pricing disability income insurance. Those techniques for which claim costs can be used and those for which the pricing actuary must project cash flows are identified.

Key words and phrases: benefits, claim costs, cash flows, incidence rate, exposure, survivorship function

1 Purpose and Product Features

1.1 Purpose

One of the primary purposes of life insurance is to protect a family from the catastrophic financial consequences of a breadwinner's premature death. Death is not the only reason, however, that a breadwinner may be unable to work. For example, what if the breadwinner is a victim of serious illness or injury and is unable to work for a significant period of time? Will his or her dependents be worse off financially than if he or she had died?

*Mark Chartier received his B.Sc. in mathematics from the Massachusetts Institute of Technology in 1980. He served four years in the United States Navy before joining Monarch Life Insurance Company. Afflicted with a congenital inability to pass fellowship exams, he bears the title "Career Associate" as a badge of honor.

Mr. Chartier's address is: Monarch Life Insurance Company, One Monarch Place, Springfield MA 01133, USA.
This other need defines the purpose of disability income insurance: to provide income when illness or injury renders the insured unable to work. The need is all too real and is arguably greater than that for life insurance. For example, according to the 1980 Commissioner's Standard Ordinary Table, a male age 35 has a 0.00211 probability of dying during the next year. According to the 1985 Commissioner's Individual Disability Tables A, the probability that a 35 year old male will suffer a disability lasting at least 90 days at some time during the next year ranges from 0.00164 in the best occupation class to 0.01176 in the worst occupation class.

Most United States residents have two forms of disability income protection provided by social insurance: workers' compensation and Social Security. Although workers' compensation is technically liability insurance for an employer, it effectively provides disability income protection to the employee. There are, however, some important limitations to workers' compensation. An obvious one is that it only protects employees from disabilities that are work-related. Another is that some segments of the population (the self-employed, for example) are not covered.

Social Security provides much broader protection. The \(D\) in OASDI represents the United States' most comprehensive response to the need to care for the incapacitated. There are, however, serious shortcomings to Social Security including: (i) a restrictive definition of disability that makes it difficult to qualify for benefits; (ii) a claim settlement process that does not determine eligibility for benefits on a consistent basis; and (iii) an absence of promises. Entitlement to Social Security benefits of any kind is a statutory right (not a contractual right) and the terms of coverage can be changed at any time by an act of Congress. No promise is made about the level of benefits, the size or the timing of any cost of living adjustment, or the definition of disability used to determine eligibility for benefits. Only private insurers can make promises by entering into contracts with private individuals. See Rejda (1984, Chapter 2, pp. 19-46) for a more detailed discussion of the shortcomings of Social Security.

The providers of individual disability income insurance form a small segment of the insurance industry. Based on net earned premium fig-

---


\(^3\)*Old-Age, Survivors, Disability and Health Insurance* (OASDI) refers to the monthly retirement, disability, and survivor benefits paid under the United States Social Security system.
ures for 1992, more than half of the business is written by only four carriers: Paul Revere, Provident Life and Accident, Northwestern Mutual, and UNUM. (See Conning & Company, 1993.)

Similarly, not many Americans purchase individual disability income insurance. According to Conning & Company (1993), 75 percent of the individual disability income insurance in-force is on the top occupation class, the professional elite consisting mainly of physicians, attorneys, highly paid corporate executives, certified public accountants, actuaries, etc. For example, Soule (1993), reports that 80 percent of dentists, 78 percent of physicians, and 68 percent of lawyers have disability income insurance. The situation with group disability income insurance is slightly better. Contrary to popular belief, however, most Americans do not have group long-term disability income protection. According to Goldman (1990), while there were over 50 million workers covered by short-term disability (benefit period of two years or less) in 1986, the number covered by group long-term disability was small, less than 20 million.

1.2 Policy Features

Below is a list of some of the more common policy features of individual disability income insurance. This is by no means an exhaustive description. The reader should consult Kidwell (1988, Chapter 3) for more detailed information.

1. Rate Structure: Premiums can be level (in which case they vary by issue age but not by attained age) or they can increase by attained age. Level premium rate structures are common and involve a significant prefunding of future benefits. This prefunding is quantified in the active life reserve (to be discussed later).

2. Renewability: Three categories are common: nonrenewable for stated reasons only, guaranteed renewable, and noncancellable.

A contract is nonrenewable for stated reasons only if the insurer reserves the right to cancel coverage but may do so only for one of the reasons stated in the contract. For example, the contract might specify that the insured must be working full time in order to renew. The premium rate the insured pays is not guaranteed for the life of the contract.

---

4Group disability income coverage is a different topic. This important subject will not be discussed here. See Goldman (1990) for more on group disability income insurance.
A contract is *guaranteed renewable* if the insurer is contractually prohibited from canceling coverage for any reason other than failure to pay premiums. The premium rate the insured pays is not guaranteed. Even when the contract has a level rate structure, the insurer reserves the right to increase the rate if experience is significantly worse than anticipated when the contract originally was priced.

The most generous renewability provision is found in *noncancellable contracts*. The insurer makes timely payment of premiums the only condition for renewal and guarantees the premium rate until policy expiration, typically at age 65. An insured who buys such a contract with a level rate structure at age 25 will pay the same rate for his or her coverage for the next 40 years.

3. **Elimination Period**: The elimination period is the period of time for which the policyholder self-insures. The insured only begins to accrue benefits after the elimination period is completed. The elimination period can be likened to the deductible on a medical expense policy. Let's illustrate the concept by a few examples.

Suppose the insured has a contract that, should he or she become disabled, will pay an annuity of $100 per month. Furthermore, assume the insured is disabled for 45 days before returning to work:

(a) If the insured's contract has a zero day elimination period, he or she is paid $150 (one and a half month's benefit) for those 45 days.

(b) If the insured's contract has a 30 day elimination period, he or she would be paid $50 (half a month's benefit) for the remaining 15 days.

(c) If the insured's contract has a 60 day elimination period, he or she would accrue no benefit because the disability did not extend beyond the end of the 60 day elimination period.

This paper distinguishes time on claim and time disabled. A claim does not begin until the elimination period is completed. The length of time on claim plus the length of the elimination period equals the length of time disabled. Because the longer the elimination period, the lower the premium the insured must pay, the choice of an elimination period is important. It should reflect the

---

5 One month is assumed to have 30 days
insured's judgment of how long he or she can rely on personal savings. Elimination periods of 30, 60, and 90 days are common. For financially well-endowed individuals, 180 day and 365 day elimination periods are available.

4. Benefit Period: This is the maximum length of time an insured can collect benefits for a single claim. Benefit periods can range from a few months to the entire life of the insured. A common benefit period is one that would allow the policyholder to remain on claim until age 65.

5. Definition of Disability: Determining whether to pay a life insurance benefit is a relatively straightforward matter. This is not the case for disability income insurance; the potential for policyholder abuse is enormous. It is essential that the policy state as precisely as possible what constitutes a disability. Two definitions are common. The most generous is the regular occupation definition. According to this definition, the insured is disabled if an illness or injury renders her or him unable to perform the substantial and material duties of her or his regular occupation; whether the insured is able to perform the duties of some other occupation is irrelevant.

An alternative is the reasonable occupation definition under which the insured is disabled if he or she is unable to do the substantial and material duties of any reasonable occupation. A reasonable occupation is one the insured could be expected to perform by virtue of education, training, and work experience. Some insurers modify the definition of a reasonable occupation by asserting that due regard must be given to the insured's earnings before disability began. The intent is to protect the insured from being forced into an occupation in which he or she would suffer a substantial loss of income.

Let us illustrate the difference between these two definitions with a realistic example. Suppose an insured's regular occupation is that of a surgeon. The insured begins to suffer from an impairment in her or his right wrist, perhaps carpal tunnel syndrome or arthritis. He or she no longer can perform surgery. But by entering general medical practice, the surgeon still can earn a much higher income than that of the average individual. Is the surgeon disabled? By the reasonable occupation definition, the answer is no; the surgeon now can perform the duties of what is for her or him a reasonable occupation. By the regular occupation defini-
tion, however, the surgeon is considered to be disabled because he or she cannot perform the duties of his or her regular occupation. Therefore, the surgeon can collect benefits while she or he earns income from a new occupation!

Some policies offer a hybrid of the two definitions. For example, the regular occupation definition may apply only during the first 24 months of a claim. After 24 months, a claimant can continue to collect benefits only if he or she meets the reasonable occupation definition.

6. Partial or Residual Disability: The above definitions refer to total disability, the complete inability to perform the substantial and material duties of the insured's regular or some other reasonable occupation. What if the disabled insured can work, but only on a part-time basis? What if she or he can work full-time but can perform only some of the key duties of the job? If a contract only covers total disability, he or she is not eligible for benefits. Some contracts will pay a fraction of the policy's full benefit amount under a partial or residual disability clause. Under a partial disability clause, the benefit is a function of time unable to work. Under a residual disability clause, the benefit is a function of income lost. Contracts providing residual benefits are common and will be discussed later.

7. Presumptive Disability: Some contracts will assert that for certain conditions, the policyholder will be presumed disabled and able to collect benefits even if he or she continues to work and suffers no loss of income. Such conditions might include total blindness in both eyes, loss of use of both hands or both feet, total deafness, etc. The presumptive disability provision also might extend the benefit period under such circumstances, e.g., pay benefits for life in case of blindness.

8. Protection Against Overinsurance: The insured is not supposed to profit from a disability claim. Therefore, the insurer should not issue so much coverage that when social insurance and any other private insurance is added, the insured's income on claim is higher than the income before disability. There are policy provisions that can be added to a disability contract to provide the insurer with added protection against overpayment.

One example is a coordination with social insurance clause. Under such a clause, the amount the insured is paid by his or her
private insurer can be reduced dollar for dollar by what he or she receives from social insurance. If a contract has a benefit amount of $1000 per month and a claimant receives $400 a month from OASDI, then he or she will receive only $600 per month from private insurance. Other variations are possible. The formula for coordination of benefits need not be a dollar-for-dollar offset. It may be a percentage reduction in the event social insurance is received.

The above provision applies only to social insurance. It does not take into account other private insurance policies the insured may own. A more comprehensive provision is the relation of earnings to insurance clause, an optional provision under the NAIC's Uniform Individual Accident and Sickness Policy Provision Law. For a discussion of the Uniform Law, see O'Grady (1988, Appendix 2). Such a clause allows the benefit the insurer would pay to be reduced proportionately. Suppose that the insured's income before disability is only 90 percent of the sum of all of his or her disability income insurance benefits. If a contract contained the relation of earnings to insurance clause, the insurer would pay only 90 percent of the policy's benefit amount. Neither of the above provisions typically is found in contracts sold to the industry's target market, the self-employed professional; intense competition has pushed the industry to produce very generous contracts.

9. **Waiver of Premiums:** The typical waiver of premium provision specifies that once a disabled insured has satisfied a certain waiting period (typically longer than the elimination period), premiums will be waived for the duration of disability and any premium paid after commencement of disability will be reimbursed.

10. **Riders:** Various optional benefits can be purchased to supplement the basic contract. A rider can be purchased to adjust benefits for cost of living increases while the insured is on claim. The insured can pay for the right to purchase additional coverage in the future without evidence of medical insurability. The insured could purchase a social insurance contingency rider; if he or she applies for Social Security and is turned down, he or she can collect benefits under this rider. Halpern (1979) discusses how to price such a rider.

---

The National Association of Insurance Commissioners (NAIC) is an association consisting of state insurance commissioners. The NAIC drafts model laws and recommends their adoption by state legislatures. The NAIC has no legal authority to force states to enact its recommendations.
1.3 Risk Variables

Claim experience is affected by many variables including the following: age, gender, elimination period, benefit period, relationship of benefit amount to income, occupation class, and state (geographical location) of residence. The effects of some of these variables are quantified in the rate manual. For example, premium rates increase with issue age and benefit period, and they decrease with elimination period. Some insurers use a single rate manual for the entire nation, others have surcharges in high risk states and discounts in low risk states. Women, especially those in the child-bearing ages, tend to be charged a higher rate than men, a practice justified by higher claim incidence rates. Occupation class is a crucial risk variable in disability income insurance. As a life insurance risk, a carpenter may be no different from an office worker; however, this is not the case in disability income insurance. Disability insurers therefore group occupations into broad risk classes. The class charged the lowest premium rate normally contains highly skilled professionals: physicians, lawyers, accountants, etc. The class charged the highest premium rate normally contains occupations involving substantial manual labor.

Unfortunately, the risk variable that is the hardest to measure (and is probably the most significant) is the motivation to work. A highly motivated insured with some health problems may be a better risk than a healthy insured who is willing to turn a questionable condition into a claim. The problems associated with malingering make it more difficult for the underwriter to judge the disability risk than to judge the life risk. It is crucial that no applicant be allowed to purchase a higher benefit amount than is justified by his or her income.

1.4 Overview

We will explore some of the mathematics of disability income insurance in the U.S. and Canada. In particular, Section 2 describes the underlying model and the basic notation used throughout the paper. Section 3 describes the concept of claims costs and contrasts it with benefit cash flows. The statutory active life reserve and the disabled life reserve are discussed in Section 4. Section 5 describes two approaches to calculating profits: the claim cost profit model and the statutory profit model. Section 6 provides an overview of the techniques used in pricing disability income insurance: the loss ratio technique, the percentage of premium profit method, and the asset share and the return on investment techniques. Section 7 shows how to calculate cash flows. Section
8 deals with waiver of premiums. Section 9 investigates the impact of relaxing some of the assumptions (in Section 2) on the model. Section 10 reviews a sample asset share calculation presented by Bluhm and Koppel (1988, Chapter 4). The appendix contains a numerical example.

For an overview of the mathematics of disability income insurance in Europe, see Gregorius (1993), Hertzman (1993), Mackay (1993), and Segerer (1993).

2 The Model

Consider a closed block of individual disability income insurance policies (to be described in Section 2.1). This block is assumed to consist of homogeneous business cells in which all policyholders have common parameters (characteristics) such as issue age, occupation class, elimination period, waiting period for waiver of premiums, and gender. Each cell is defined by the value of \( \theta \), a vector of parameters. To be precise, \( \theta = (\text{issue age}, \text{gender}, \text{occupation class}, \text{elimination period}, \text{waiting period for waiver of premiums}) \). For example, a cell may consist of policies sold to female surgeons with issue age 45, a 30 day elimination period, and a 180 day waiting period for waiver of premiums. This yields \( \theta = (45, \text{female}, \text{surgeon}, 30, 180) \).

Throughout the rest of this paper, the symbol \( \theta \) will be used to denote a particular business cell. We will develop functions and expressions that are dependant on \( \theta \). In most of these cases, \( \theta \) appears as a subscript.

2.1 The Policy

Consider an individual disability income insurance policy with the following features:

1. The policy is sold only to those individuals who are active (not disabled) at the time of issue.

2. An insured receives benefits if and only if the insured is disabled for a continuous period of time in excess of the elimination period. The definition of disability is not important here, suffice to say that the definition chosen will affect the probabilities of disability and recovery.

3. Premiums are not necessarily level and are paid annually.
4. There is a waiver of premium if the insured remains disabled for a period of time in excess of the waiting period for waiver of premiums. Premium payments resume upon recovery.

5. Benefits are paid up to age 65.

6. The unit of benefit is $1/12 per month.

2.2 Mortality, Morbidity, Recovery, and Lapses

There are three sources of decrement at work on the in-force population: (i) voluntary lapsation; (ii) death while on claim; and (iii) death while not on claim. The mortality rates of insureds on claim are different from those of insureds not on claim. In addition, the mortality rates of insureds who have been on claim and who since have recovered may be different from those who have never been on claim. Finally, the voluntary lapse rates of insureds who have been on claim are almost certainly lower than the voluntary lapse rates of insureds who have never been on claim. Those who have benefited from the contract are more likely to hold onto it.

In practice, the disability income insurance actuary uses a single set of tables, called lapse tables, to decrement the in-force population. These tables express the three sources of decrement mentioned above as a single aggregate source of decrement. Separate lapse tables for insureds who have been on claim and insureds who have never been on claim are not used.

The in-force population can be divided into two subpopulations: the active population and the claim population.

1. The active population is the population exposed to the risk of disablement. There are three sources of decrement and one source of increment on the active population:

(a) Voluntary lapsation;
(b) Death while not on claim;
(c) Migration from the active population to the claim population; and
(d) Recovery from claim (the one source of increment).

We will assume initially that we can identify the active population with the in-force population. That is, the population exposed to risk will be calculated by taking the in-force population at time of issue and decrementing only for voluntary lapsation, death on claim, and death not on claim.
2. The claim population consists of those persons who are receiving disability benefits. Thus, disabled persons who are not receiving a benefit are not considered as being on claim. There are two sources of decrement and one source of increment on the claim population:

(a) Death while on claim;
(b) Recovery from claim; and
(c) Going on claim (the one source of increment).

When measuring the size of the active population during the year, migration to the claim population and return from claim to the active population will be ignored. We make this assumption for two reasons. First, the claim population is very small relative to the active population, so little error results from identifying the active population with the entire population in-force. Second, tracking the continuous two way migration between active and claim population is a tedious process, and we want to keep the model as simple as possible. In Section 9.2, we will describe a way to track this two way migration at discrete intervals.

A set of morbidity tables with two decrements (death on claim and recovery from claim) is used to project increments and decrement to the claim population. Claim incurral is the sole source of increment. As with lapse rates, the disability income insurance actuary does not use a multiple decrement table. Instead he or she uses claim termination rates that combine termination due to death and recovery.

The rate of claim incurral for insureds who have never filed a claim is undoubtedly different from the incurral rate for those who have filed claims in the past. In practice, the disability income actuary uses aggregate rates of claim incurral drawn from a single set of morbidity tables that do not distinguish the two groups of insureds. In addition, because policyholders are assumed to pay premiums annually, voluntary lapsations only occur at policy anniversaries. While premiums are being waived for a disabled insured, the policy cannot lapse.

Given these three decrements (mortality, morbidity, and lapses) and the increment (recovery), a separate combined mortality-disability-lapse table is used for each occupation class and gender. The mortality-recovery table for disabled lives will be a select table with age at time of disability and duration of disability as the parameters in the table.

For ease of computation, the following assumptions are made:

1. The incidence of claim is distributed uniformly throughout the year. Thus, given that there are $E_\theta(n)$ units of claim at the start of the $n$th policy year and that $r_\theta(n)$ is the incidence rate of claim,
then the expected number of claims in the interval \((t, t + dt)\), with \(0 < t < 1\), is \(E_\theta(n)r_\theta(n)dt\).

2. Deaths and lapses in the in-force population occur at the end of the policy year.

3. The population on claim is so small relative to the active population that we initially will assume that the units of insurance exposed to risk are unaffected by claim incursals and recoveries. The effects of relaxing these assumptions are investigated in Section 9.

2.3 Notation

Some of the more basic symbols used in the model will be defined and assumptions for the model presented.

\[ \theta = \text{A vector of parameters that characterize each business cell;} \]
\[ i = \text{The valuation rate of interest;} \]
\[ \nu = \text{The annual discount factor, i.e., } \nu = 1/(1 + i); \]
\[ x = \text{The issue age, } x = 15, 16, 17, \ldots; \]
\[ z = \text{The attained age } = x + n - 1; \]
\[ n = \text{The policy year of disablement, } n = 1, 2, 3, \ldots; \]
\[ m = \text{The current policy year } m = n, n + 1, \ldots; \]
\[ e = \text{The length of the elimination period (measured in years). It is the minimum length of time the insured must be disabled in order to qualify for benefits;} \]
\[ w = \text{The length of the waiting period (measured in years) to qualify for waiver of premium, with } w > e; \]
\[ b = \text{The length of the benefit period (measured in years). Usually the benefit period extends to age 65. Note, } b > w > e; \]
\[ r_\theta(n) = \text{The incidence rate of claim in the } n\text{th policy year for a policy with parameter } \theta, \text{ i.e., the probability that an active insured (currently age } x + n - 1\text{) from the business cell with parameter } \theta \text{ becomes disabled and remains disabled for at least the length of the elimination period;} \]
\( E_\theta(n) = \) The number of units\(^7\) of insurance exposed to risk in the \( n \)th policy year for a policy with parameter \( \theta \);

\( P_\theta(n) = \) The annual premium per unit of insurance in-force during the \( n \)th policy year for a policy with parameter \( \theta \);

\( s_\theta(y, n) = \) The probability that a policyholder (with parameter \( \theta \)) who is disabled in the \( n \)th policy year and remained disabled throughout the elimination period will stay on claim (i.e., receive benefits) for at least \( y \) consecutive years into the future. Also, \( s_\theta(y, n) = 0 \) when \( y \) is greater than the length of the benefit period;

\( BC_\theta(n, m) = \) The benefit cash outflow in the \( m \)th policy year to claimants (with parameter \( \theta \)) who are disabled in the \( n \)th policy year; and

\( W_\theta(n, m) = \) The amount of premium waived at the beginning of the \( m \)th policy year on claimants (with parameter \( \theta \)) who are disabled in the \( n \)th policy year.

### 3 Claim Cost

Before introducing the concept of a claim cost, let us review the way \( PD_B \theta(n) \), the projected death benefit cash outflow in the \( n \)th policy year on a single life insurance policy with parameter \( \theta \), is calculated. Once mortality and lapse assumptions are chosen for pricing purposes, the life actuary easily can determine \( PD_B \theta(n) \) as follows:

\[
PD_B \theta(n) = q_\theta(n)DB_\theta(n)
\]

where \( q_\theta(n) \) is the mortality rate in the \( n \)th policy year, and \( DB_\theta(n) \) is the number of dollars of death benefit exposed to risk during the \( n \)th policy year for a policy with characteristics \( \theta \).

For a disability income product, however, the calculation is more complex because the benefit is an annuity, not a single lump sum payment. The pricing actuary simplifies the problem by calculating a claim cost. For policies issued at age \( x \), the claim cost associated with the \( n \)th policy year, \( CC_\theta(n) \), is given by:

\[
CC_\theta(n) = v^e r_\theta(n)E_\theta(n)a_\theta(z)
\]

\(^7\)A unit of insurance is defined as \$1 of annual benefit paid monthly until recovery or the end of the benefit period.
where \( a_\theta(x) \) is the actuarial present value of an annuity paying \$1 per year (paid monthly) after the elimination period ends to a life who is disabled at age \( z \) in the \( n \)th policy year.\(^8\) In other words, a claim cost associates with a policy year the present value of all future monies that will be paid to a policyholder whose date of disablement is in that year.

In life insurance, future death benefits in a closed group of cohorts are a function of two random processes: lapses and mortality. In disability income insurance, the extent of future claim payments in a closed group of cohorts depends on lapses, the incidence of claims, and the severity of claims. The fact that the claim severity is random is one of the reasons why the claim experience on disability income insurance is inherently more volatile than the claim experience on life insurance. It is also one of the reasons why disability income insurance has more in common with property/casualty insurance than it has with life insurance.

Lapses, the incidence of claims, and the severity of claims vary by age, gender, occupation class, elimination period, policy age, and the contract's definition of disability. They also are affected by many other factors, some of which the pricing actuary cannot quantify easily (such as the state of the economy, for example). The claim incidence rate increases with age. In addition, the claim incidence rate decreases as the elimination period increases; it is higher for women than for men, at least when women are in their child-bearing years—it is possible this relationship reverses at advanced ages. The greater the physical stress of an occupation, the higher is the claim incidence rate.

The severity of a claim (i.e., the annuity factor) also varies with the benefit period and the interest rate. The seriousness of a claim generally increases with age. On the other hand, the benefit period shrinks as the insured ages. A policyholder with a benefit period to age 65, for example, can stay on claim for ten years if disabled at age 55. He or she can stay on claim only for five years if disabled at age 60. This can lead to a curious pattern: as a block of policies ages, claim costs first will increase. As insureds reach their late fifties or early sixties, claim costs can decrease as the shrinking benefit period causes claim severity to become smaller.

Unlike claim incidence, claim severity rises as the elimination period increases. This is because a long elimination period screens those claims that would have closed relatively early. For example, in comparing an insured with a 30 day elimination period to an insured with a 90 day elimination period, the latter must be disabled more seriously.

\(^8\)By convention, \( CC_\theta(n) \) is valued as of the date of disablement.
than the former to get on claim. But if they get on claim, the latter is expected to remain on claim longer. The author frequently has observed that claim severity in the less skilled occupation classes is less than that in the highly skilled occupation classes. In other words, blue and gray collar workers return to work faster than white collar workers. This is the case even when differences in elimination period and age distribution are accounted for.\textsuperscript{9}

Let $H_\theta(z)$ denote the claim cost per unit at attained age $z$, then:

\[
H_\theta(z) = \nu^e r_\theta(n) a_\theta([z + 1/2] + e, b)
\]

where $a_\theta([z + 1/2] + e, b) =$ the actuarial present value of an annuity of $1$ per year (paid monthly) for at most $b$ years starting at age $z + 1/2 + e$ to a life (with parameter $\theta$) disabled at age $z + 1/2$. Age at disability is taken to be $z + 1/2$ because it is assumed that, on average, disability begins in the middle of the policy year. The annuity starts at $z + 1/2 + e$ because payment commences after the elimination period is completed. The annuity is contingent on the insured remaining disabled. The claim cost now can be rewritten as

\[
CC_\theta(n) = E_\theta(n)H_\theta(z).
\]

4 Reserves

First let us introduce two important items that appear on the balance sheet of the disability income insurer: the active life reserve and the claim reserve. Roughly speaking, the active life reserve is that part of the liability for future claims (yet to occur) that must be prefunded, and the claim reserve is the liability for claims that already have been incurred. In more precise language, the active life reserve is the expected present value of future claim costs minus the expected present value of future net premiums. It is analogous to the policy reserves held on life insurance contracts. The claim reserve is another item that makes disability income insurance similar to property/casualty insurance. It is the expected present value of all future payments, both contingent and noncontingent, that will be made on claims that have been incurred. For a more thorough discussion of the active life reserve and the claim reserve, see Bartleson (1968) and Shapland (1988, Chapter 5).

\textsuperscript{9}One explanation may be that the higher claim incidence rate among blue collar workers causes the lower severity. High claim incidence means there are many claims for conditions that are not serious, so they close quickly. Another is that insureds in the lower skilled occupations have less generous contracts, giving them greater incentive to return to work.
The formula for the statutory active life reserve at the beginning of the nth policy year, $V_{\theta}^{(aa)}(x, n)$, on a single unit of insurance issued at age $x$ and continuable until attainment of age 65 is:

$$V_{\theta}^{(aa)}(x, n) = \sum_{k=0}^{65-z} \nu^k k p_z^{(aa)} \left[ \nu^{1/2} p_{z+k}^{(aa)} H_\theta(z+k) - P_\theta^* \right]$$

(3)

where $H_\theta$ is defined in equation (1), and $k p_z^{(aa)}$ is the probability that an active life age in cell $\theta$ (with issue age $x$) survives $k$ years.\(^\text{10}\) To avoid cumbersome notation, the $\theta$ is not shown; $P_\theta^*$ is the statutory net level premium payable to age 65 for a unit of insurance issued an active life age in cell $\theta$.

Again, a claim is assumed to be incurred, on average, in the middle of the policy year; hence, the presence of $1/2$ in the exponent of $\nu$ and the subscript of the probability of survival $p$. We assume that the reserve is established on a level premium contract. The terminal active life reserve is calculated by multiplying the above reserve per unit by the number of units in-force. As equation (3) is to be used to calculate statutory reserves, values of $H$, $\nu$, $p$, and $P$ will be specified by state regulation and may bear little resemblance to the actuary's best guess assumptions about interest rates and future morbidity.

The claim reserve is more complex. At any given time, the population of insureds on claim will be the sum of several closed cohorts. Each cohort is defined by the amount of time it has been on claim. Let $R_\theta(n, m)$ be the statutory claim reserve at the beginning of the mth policy year (or the end of the $(m-1)$st policy year) on those claimants disabled during the nth policy year. The total claim reserve at the beginning of the mth policy year, $R_\theta(m)$, is the sum of the claim reserve on $m-1$ cohorts,

$$R_\theta(m) = \sum_{n=1}^{m-1} R_\theta(n, m).$$

Now $R_\theta(n, m)$ is a function of two quantities: (i) the number of units of insurance disabled in the nth year still on claim at the beginning of the mth year; and (ii) the claim reserve established on a single unit disabled in the nth year and still on claim at the beginning of the mth year. Later we will develop the formula for the number of units disabled in the nth year still on claim at the beginning of the mth year. The claim

\(^{10}\)Once a person is insured, a reserve is maintained for her or him. Even if she or he is on claim, an active life reserve is maintained for her or him because she or he could go on claim again after recovery.
reserve factor for a single unit of insurance disabled in the \( n \)th year and still on claim at the beginning of the \( m \)th year is \( V^{(ii)}_{\theta}(n, m) \), i.e.,

\[
V^{(ii)}_{\theta}(n, m) = \frac{12(b-(m-n-e-1/2))}{12} \sum_{k=1}^{12} v^{k/12} \frac{1}{12} p^{(ii)}_{k/12(z+1/2)+m-n-1/2} \tag{4}
\]

which is the present value of a $1/12 per month annuity. The notation \( k/2 P^{(ii)}_{z+1/2+t} \) denotes the probability an insured who became disabled at exact age \( z+1/2 \) and has remained disabled for \( t \) years will remain on claim for at least another \( k \) months. The term \( 12(b-(m-n-1/2-e)) \) is the number of months remaining in the benefit period. Again, we assume that disablement occurs on average in the middle of the policy year. The values of \( p^{(ii)} \) and \( v \) are specified by statutory regulation and may not coincide with the pricing actuary’s best guess estimate of what will happen.

Let \( D_{\theta}(n, m) \) be the number of units of insurance disabled in the \( n \)th year and still on claim at the start of the \( m \)th policy year; then

\[
R_{\theta}(n, m) = V^{(ii)}_{\theta}(n, m) D_{\theta}(n, m). \tag{5}
\]

5 Profits

We present here two general approaches to measuring profit (for a homogeneous group of policies with parameter \( \theta \)) in disability income insurance. Each provides a formula for annual recognition of profit. It should be evident by the end of this discussion that the two approaches do not recognize the same profit year by year. Throughout this section, we will drop the subscript from the symbols to reduce clutter. We must remember, however, that the totals in this section apply to the business cell with parameter \( \theta \).

5.1 The Claim Cost Profit Model

Under a claim cost pricing model, the total profit from cell \( \theta \) in the \( n \)th policy year, \( PROF^{(cc)}_n \), is given by the following formula:

\[
PROF^{(cc)}_n = P_n + I_n - EXP_n - COM_n - CC_n - (ALR_n^{(e)} - ALR_n^{(b)}) \tag{6}
\]

where \( P_n \) is the total premium earned; \( I_n \) is the total investment income; \( EXP_n \) is the total expenses; \( COM_n \) is the total commissions from cell \( \theta \); \( CC_n \) is the total claim cost incurred during the \( n \)th policy year; \( ALR_n^{(b)} \) is
the total active life reserve at the end of the year; and $ALR_n^{(b)}$ is the total active life reserve at the beginning of the $n$th policy year. Premiums, expenses, and commissions are calculated in much the same way as they are for life insurance. The $ALR$ terms are calculated by multiplying equation (3) by the number of units of insurance in-force, i.e.,

$$ALR_n^{(b)} = V_\theta^{(a)}(n)E_\theta(n)$$
$$ALR_n^{(e)} = V_\theta^{(a)}(n+1)E_\theta(n+1).$$

We uncover a serious flaw in the claim cost pricing model when we try to calculate investment income, $I_n$. The reader may be tempted to calculate it as:

$$I_n = i \times (P_n + ALR_n^{(b)} - EXP_n - COM_n - \nu^{1/2}CC_n)$$

where $i$ is the assumed rate of interest. Again, we have assumed that claim costs are incurred in the middle of the policy year. It is a mistake to subtract the entire value $\nu^{1/2}CC_n$, however, because $CC_n$ is a lump sum representing a series of cash flows that may be spread over many future years. Only the portion that is disbursed in the current year should be subtracted; the insurer is free to invest the remainder until payment is due. Because the claim cost pricing model does not divide that lump sum into money disbursed now and money disbursed later, it cannot correctly allocate investment income by policy year. The statutory profit model corrects this flaw.

5.2 The Statutory Profit Model

Statutory book profit, $PROF_n^{(s)}$, is given by the equation:

$$PROF_n^{(s)} = P_n + I_n - EXP_n - COM_n - BEN_n$$
$$- (CR_n^{(e)} - CR_n^{(b)}) - (ALR_n^{(e)} - ALR_n^{(b)})$$

where $BEN_n$ represents the actual benefits paid in cell $\theta$, $CR_n^{(b)}$ and $CR_n^{(e)}$ are the total claim reserve at the beginning and the end of the $n$th policy year. The only apparent difference between equations (6) and (8) is that claim cost has been replaced by benefits paid plus the change in claim reserve. Are equation (6) and equation (8) equal? They typically will not be. Claim costs are projected using pricing mortality-morbidity tables and interest rates. Statutory claim reserves are measured using statutory mortality-morbidity tables and interest rates. It is highly unlikely that quantities derived from different assumptions will be equal.
Under the statutory profit model, investment income is:

\[ I_n = i \times (P_n + ALR_n^{(b)} + CR_n^{(b)} - EXP_n - COM_n - \nu^{1/2} BEN_n). \]  

(9)

Notice that in equation (9) investment income is counted on all monies held in reserve and that the active life reserve earns interest, but the claim reserve is ignored. In equation (9), benefits paid in a given year are subtracted from revenue before applying an interest rate, but money for future benefits earns interest until the benefits are paid. If the pricing actuary needs to project the year by year pattern by which statutory profit emerges, claim costs cannot be used. The pricing actuary instead must calculate benefit cash flows.

6 Overview of Pricing Techniques

This section discusses some of the methods used to determine premium rates for disability income insurance. The list is by no means exhaustive.

6.1 The Loss Ratio Technique

Before describing this technique, we must define what is meant by the loss ratio. The term loss ratio usually is understood to be the fraction of the policyholders' premiums that is returned in benefits. That sounds simple enough, yet there is great confusion about what a loss ratio is and significant disagreement about how it is calculated. A loss ratio can be retrospective or prospective. It can be applied over the entire life of a block of business or to a single experience year. It can be calculated with GAAP reserves, statutory reserves, natural reserves, or no reserves at all. It can be calculated using a realistic interest rate, a statutory interest rate, or no interest rate at all. Claim settlement expenses can be added to the numerator of the loss ratio (property/casualty insurance) or play no part in the calculation of the loss ratio (health insurance). (See Pharr, 1979, with discussion.)

For our purposes, the term loss ratio is defined to be the ratio of the expected present value of future benefits that will be paid over the business cell to the expected present value of future premiums that will be collected from the business cell. This is a prospective lifetime loss ratio. When calculating the relevant present values, the author recommends that we discount for voluntary lapsation as well as mortality and that we use a realistic rate of interest, not the statutory valuation rate. This view is not shared by all actuaries.
Once the question of how to calculate a loss ratio is settled, the loss ratio technique becomes the simplest approach of all. It can be broken down into three steps:

1. Calculate the present value of future claim costs on a single unit of insurance.

2. Calculate a level net premium rate by dividing the result of step 1 by an annuity factor.

3. Calculate a gross premium by dividing the net premium from step 2 by a target loss ratio. For example, assume the present value of future claim costs is $50. Assume the present value of an annuity due of $1 per year over the life of the insurance contract is $5. Assume the target loss ratio is 50 percent. Then $50/(5\times0.5) = $20 is that gross premium rate for which the expected present value of future benefits divided by the expected present value of future premiums is 50 percent. Note that in step 1 you could calculate the present value of future benefit cash flows in place of future claim costs. Claim costs, however, are suited perfectly for this technique.

There are advantages and disadvantages to the loss ratio method. The most important advantage is its simplicity. The formulae are easy to understand, and no assumptions for future expenses are needed. One doesn't even need to know precisely what the commission scales are. This simplicity is also its most important disadvantage; the difference between gross and net premiums may not be sufficient to cover expenses and commissions. Nevertheless, if a company's expenses or commissions are unacceptably high, the pricing actuary may be compelled to use the loss ratio technique.

A minimum loss ratio is required by law or regulation in most states. Taking the minimum loss ratio as a pricing target, the actuary can solve for the maximum premium that legally can be charged. Once the actuary knows how much premium is left after paying benefits, he or she can solve for target expense and commission levels down to which actual levels should be managed. Another advantage is that if the target loss ratio is set high enough, the method automatically produces a rate manual whose anticipated loss ratio exceeds the required minimum. This has not been a significant advantage of late. Anticipating a certain loss ratio is one thing, experiencing it is another. Despite all the confusion regarding correct loss ratio calculation, one unambiguous lesson has emerged from the last half of the 1980s: the least of the industry's worries is that the loss ratio will be too low.
6.2 Percentage of Premium Profit Method

This method is similar to the equation method in life insurance except that the present value of future death benefits is replaced by the present value of future claim costs. The pricing actuary takes as his or her target a certain percentage of the present value of premiums that will go to profit. The actuary projects the present value of claim costs, expenses, etc., and solves for that premium rate at which a sufficient percentage of premium will be left to meet or exceed the target set. Claim costs are suited perfectly to this technique. Again, the present value of future claim costs could be replaced by the present value of future benefit cash flows to obtain the same result.

Because the percentage of premium profit method makes explicit provision for commissions, expenses, and taxes, it is superior to the loss ratio technique. A major shortcoming of the percentage of premium profit method is that it does not quantify the risk/reward relationship. How great a percentage of premium profit is sufficient to compensate the insurer for taking the disability risk? How does a risky disability income portfolio with a 10 percent of premium profit compare with a risk-free Treasury bill paying a 3 percent return?

As noted earlier, disability income insurance has volatile claim experience. In part this is due to the extra random process in morbidity, the claim severity, and the fact that the insured exercises some control over morbidity, possibly electing to be on claim rather than being put on claim by forces beyond his or her control. Disability income insurance is risky business for a variety of reasons, and the faint of heart are driven from the marketplace. A pricing method should produce a rate manual that is not merely profitable, but is more profitable than one for a less risky line of business.

6.3 The Asset Share Technique

As with life insurance, the profit target for disability income insurance can be set as a certain asset share by a certain policy year. Alternatively, the target could be set as a certain level of surplus by a certain year, where *surplus* could be defined as the difference between the asset share and the total statutory reserve. Claim costs are not appropriate for the calculation of asset shares. One cannot take the asset share equation for life insurance, substitute claim cost incurred for death benefits paid, and consider the result to be the disability income insurance asset share equation.
An asset share is the accumulation, per unit of insurance in-force, of all cash inflows to date minus all cash outflows to date. Every variable in the equation for asset share represents a cash flow. A claim cost is not a cash flow, however; it is a lump sum assigned to a single policy year and is equal to the present value of a series of cash flows that may spread over several future years. In the first policy year, the claim cost is higher than the benefit cash flow because the claim cost includes payments that will be made in future years. A claim cost pricing model understates the first year asset share. There may be subsequent years in which the claim cost is less than the benefit cash flow because some claimants disabled in prior policy years will be collecting benefits in the current year; each year's claim cost measures payments only to claimants whose disabilities commenced in that year.

Bluhm and Koppel (1988) present a sample asset share calculation. An attempt is made to calculate benefit cash flows using claim costs and changes in claim reserve. After developing our own model to project benefit cash flows, we will discuss some of the problems inherent in their method and show how it can be rehabilitated.

6.4 The Return on Investment Technique

With this method, the pricing actuary projects future book profits and then solves for the return on investment (ROI). The ROI is the discount rate at which the present value of renewal year profits equals the loss in the first policy year. The profit target is a threshold ROI. In the author's opinion, this technique is superior to the other methods presented above because it quantifies the insurer's reward for bearing the significant risk of competing in the disability income marketplace. The rational investor only increases risk if he or she has a reasonable expectation of a higher return (and insurance company shareholders are, we presume, rational investors).

If the risk of selling disability income is higher than that of selling term life, then the insurer is entitled to expect a higher ROI from the former product than from the latter. The ROI pricing technique not only allows the insurer to compare disability income to other products, but allows the insurer to tailor individual rates to each particular risks that it bears. For example, the insurer bears more risk when it sells a contract with a lifetime benefit period than when it sells a contract with a benefit period of only one year. The rates should be set so that the

11 These comments are not intended to detract from the quality of the articles contained in this excellent text. The author highly recommends the O'Grady textbook.
insurer's expected ROI on lifetime contracts is higher than that on one year benefit period contracts.

It is here that we come to the most serious defect of a pricing model based on claim costs instead of cash flows. While claim costs can be used to project the present value of profit over the entire life of a block of business, they cannot be used to project the year by year pattern by which statutory book profit emerges. Consequently, a claim cost pricing model cannot measure ROI correctly. This is because claim costs are calculated with pricing assumptions, while statutory claim reserves are calculated with assumptions specified by regulatory authorities.

If pricing assumptions are more liberal than statutory assumptions (not a given in today's morbidity environment), then claim costs will be lower than benefits paid plus change in claim reserve in the early policy years. At this time a claim cost pricing model will overstate statutory book profit. In later years the inequality will reverse as money released from the claim reserve makes the real book profit higher than that predicted by a claim cost pricing model. Therefore, if pricing assumptions are more liberal than statutory valuation assumptions, a claim cost pricing model will recognize profit earlier than would emerge under statutory accounting and will overstate the ROI that will be realized.

In summary, if an insurer places any importance on estimating the asset shares or the ROI of a new disability contract it contemplates introducing, the actuary must translate pricing assumptions into cash flows, not claim costs. The remainder of this paper will be devoted to a model for doing this.

7 Cash Flows on the Base Policy

We now develop a model to calculate $BC_\theta(n, m)$, the value of the benefit cash outflow in the $m$th policy year to claimants with dates of disablement in the $n$th policy year. As an example, consider a business cell consisting of disability income policies that will expire five years after issue. Let

$$
\begin{pmatrix}
BC_\theta(1, 1) & BC_\theta(1, 2) & BC_\theta(1, 3) & BC_\theta(1, 4) & BC_\theta(1, 5) \\
BC_\theta(2, 2) & BC_\theta(2, 3) & BC_\theta(2, 4) & BC_\theta(2, 5) \\
BC_\theta(3, 3) & BC_\theta(3, 4) & BC_\theta(3, 5) \\
BC_\theta(4, 4) & BC_\theta(4, 5) \\
BC_\theta(5, 5)
\end{pmatrix}
$$

be the matrix of benefit cash flows.
The total benefit cash flow in the first year is $BC_\theta(1,1)$. In the second policy year, money is paid to some claimants with disablement dates in the first policy year, $BC_\theta(1,2)$, and money is paid to claimants with disablement dates in the second policy year, $BC_\theta(2,2)$. The total benefit cash flow in the second policy year is $BC_\theta(1,2) + BC_\theta(2,2)$, the sum of the entries in the second column of the array. Likewise, the cash flow in the third policy year is the sum of entries in the third column, and so forth for all other policy years. It is interesting to note what happens if we move across a row of the array rather than down a column. If the series of cash flows $BC_\theta(1,1), BC_\theta(1,2), BC_\theta(1,3), BC_\theta(1,4)$, and $BC_\theta(1,5)$ are discounted for interest back to policy issue, the result is the claim cost in the first policy year. The total cash outflow in the $m$th policy year, $TCO_\theta(m)$, is the sum of entries in the $n$th column, i.e.,

$$TCO_\theta(m) = \sum_{k=1}^{m} BC_\theta(k,m).$$

In addition, the total claim cost in the $n$th policy year, $TCC_\theta(n)$, is the result of discounting the entries in the $n$th row back to the $n$th policy year, i.e.,

$$TCC_\theta(n) = \sum_{m=n}^{\infty} \nu^{m-n+1/2} BC_\theta(n,m).$$

This illustrates the flexibility of a cash flow model. From cash flows you can calculate claim costs. From claim costs you cannot calculate cash flows.

Next we turn our attention to the calculation of $BC_\theta(n,m)$ for a unit of benefit. This task is divided into the three cases shown below. For simplicity, the benefit of $\$1$ per year is assumed to be paid continuously.

**Case 1, $m = n$**: $BC_\theta(n,m)$ is the benefit cash flow in the $n$th policy year to claimants with dates of disablement in that same year. $E_\theta(n)$ units of insurance are exposed to risk at the start of the $n$th policy year. The number of those units that will go on claim during the time period $(t, t + dt)$ (with $0 < t < 1$) by completing the elimination period is $r_\theta(n)E_\theta(n)dt$. At time $t + e$ these claimants will begin to accrue benefits. The number of units disabled during the time interval $[t, t + dt]$ that are still on claim at time $t + e + y$ is $r_\theta(n)E_\theta(n)s_\theta(y, n)dt$. Each unit of insurance that is on claim at time $t + e + y$ is paid $dy$ during the interval $[t + e + y, t + e + y + dy]$. The final equation is:

$$BC_\theta(n,n) = r_\theta(n)E_\theta(n) \int_{t=0}^{1-e} \int_{y=0}^{1-e-t} s_\theta(y, n)dydt. \quad (10)$$
Note the upper limits of integration. If an insured becomes disabled after time \( t = 1 - e \), he or she will not complete the elimination period before the end of the \( n \)th policy year. If an insured begins to accrue benefits at \( t + e \), then when \( y > 1 - (t + e) \), he or she will have reached the end of the policy year.

In practice, a mathematical expression for \( s_\theta(y, n) \) may not be available. Hence, we must use tabulated values from weekly or monthly claim termination rates. Therefore, equation (10) is handled best by numerical integration. An example is provided in the appendix.

**Case 2, \( m = n + 1 \)**: The calculation in this case is more complex than the previous case because there are two classes of claimants disabled in the \( n \)th policy year. One class consists of those who completed the elimination period before the end of the \( n \)th policy year. This class began to accrue benefits before the \((n + 1)\)st policy year began. The time of disablement for all members in this class is \( t < 1 - e \). The other class consists of those with time of disablement \( t > 1 - e \). Members of the latter class will not satisfy the elimination period and hence will not begin to accrue benefits until after the \((n + 1)\)st policy year has begun.

For claimants who complete the elimination period before the end of the \( n \)th policy year, the benefit cash outflow in the \((n + 1)\)st policy year is:

\[
B_{\theta}^{(1)}(n, n + 1) = r_\theta(n)E_\theta(n) \int_{t = 0}^{1 - e} \int_{y = 1 - (t + e)}^{2 - e - t} s_\theta(y, n) \, dy \, dt.
\]  

(11)

The integrand has remained the same as in Case 1, only the limits of integration have changed. As already pointed out for Case 1, a claim beginning at time \( t + e \) will have lasted for \( y = 1 - (t + e) \) years by the end of the \( n \)th policy year. Note that time 0 is the start of the \( n \)th policy year, 1 represents the end of that year, 2 represents the end of the \((n + 1)\)st policy year, and \( 2 - (t + e) \) is the time on claim for a claimant disabled at time \( t \) and persisting at least to the end of the \((n + 1)\)st year.

For those claimants disabled so late in the \( n \)th policy year they do not complete the elimination period until after the beginning of the \((n + 1)\)st year, the cash flow in the \((n + 1)\)st year is

\[
B_{\theta}^{(2)}(n, n + 1) = r_\theta(n)E_\theta(n) \int_{t = 1 - e}^{1} \int_{y = 0}^{2 - e - t} s_\theta(y, n) \, dy \, dt.
\]  

(12)

It follows that

\[
B_\theta(n, n + 1) = B_{\theta}^{(1)}(n, n + 1) + B_{\theta}^{(2)}(n, n + 1).
\]  

(13)

**Case 3, \( m > n + 1 \)**: This is the simplest case of the three. Given, as we have assumed, that the elimination period is not more than one year,
by the end of the \((n + 1)\)st policy year all claimants with disablement
dates in the \(n\)th policy year will have satisfied the elimination period.
Thus

\[
BC_\theta(n, m) = r_\theta(n)E_\theta(n) \int_{t=0}^{1} \int_{y=m-n-e-t}^{m+1-n-e-t} s_\theta(y, n) dy dt.
\]  

(14)

8 Waiver of Premium Cash Flows

Waiver of premiums can be modeled as a cash flow to the insured
where the benefit is his or her premium; that is, we assume claimants
pay their premiums and then immediately receive reimbursements from
their insurer. Because some insurers do not pay commissions on waived
premiums, the financial impact of the waiver benefit may be less than
that of 100 percent reimbursement. We assume the benefit is equiva­
 lent to 100 percent reimbursement and that the waiting period for
waiver is less than one year.

Calculating \(W_\theta(n, m)\) is easier than calculating \(BC_\theta(n, m)\). While
claim payments can occur at any time during the policy year, in our
model premiums can be waived only on policy anniversaries, making
\(W_\theta(n, n) = 0\). Thus, we only need to evaluate single integrals rather
than double integrals.

Case 1, \(m = n + 1\): \(W_\theta(n, n+1)\) is the amount of premium waived at
the beginning of the \((n + 1)\)st policy year on insureds with disablement
dates in the \(n\)th policy year. Insureds disabled in the \(n\)th policy year
divide into two classes, those disabled at time \(t < 1 - w\) and those
disabled at time \(t > 1 - w\). If an insured’s disablement occurs at time
\(t > 1 - w\), she or he will not satisfy the waiver waiting period by the time
the \((n + 1)\)st premium is due, but she or he will be reimbursed once
the waiting period is completed. If an insured’s time of disablement is
\(t < 1 - w\) and he or she still is disabled when the \((n + 1)\)st premium is
due, it will be waived.

First we will handle the case in which \(t > 1 - w\). By our assumption of
uniform distribution of claim incidence, the probability that an insured
exposed during the interval \(1 - w < t < 1\) will go on claim is \(w \times r_\theta(n)\). But \(r_\theta(n)\) merely gives the incidence rate of a disability lasting at least
e units of time. In order to go on waiver, the insured must stay disabled
an additional \(w - e\) units of time. (Recall our assumption that \(e \leq w\).)
Thus, the full probability that an insured will become disabled during
the interval \([1 - w, 1]\) and remain disabled long enough to satisfy the
waiver of premium waiting period is \(w \times r_\theta(n)s_\theta(w - e, n)\).
Those insureds who were disabled in the interval \((t, t + dt)\) with \(t < 1 - w\) will go on claim at time \(t + e\) and merely need to stay on claim for \(1 - (t + e)\) units of time (until the start of the \((n + 1)st\) policy year) to have their \((n + 1)st\) premium waived. The probability this happens is \(r_\theta(n)s_\theta(1 - e - t, n)\). Thus, the total premium waived on both of these types of insureds is:

\[
W_\theta(n, n + 1) = P_\theta r_\theta(n)E_\theta(n) [w s_\theta(w - e, n) + \int_{t=0}^{1-w} s_\theta(1 - e - t, n)dt].
\]  

(15)

**Case 2, \(m > n + 1\):** As \(w\) is not greater than one year, all claimants disabled in the \(n\)th policy year who still are disabled at the beginning of the \(m\)th policy year will be on waiver. These insureds will have remained on claim for \(m - (n + t + e)\) years. The premium waived is:

\[
W_\theta(n, m) = P_\theta r_\theta(n)E_\theta(n) \int_{t=0}^{1} s_\theta(m - n - e - t, n)dt.
\]  

(16)

In Section 2.3, we asserted that \(s_\theta(t, n) = 0\) when \(t\) is greater than the length of the benefit period. The pricing actuary must bear in mind that the benefit period for waiver of premium need not be equal to the policy's base benefit period. If periods are not equal, then the survivorship function used to project \(W_\theta(n, m)\) will be different from that used to project \(B_\theta(n, m)\).

Even a contract with a benefit period as short as one year typically will permit the insured to remain on waiver until he or she attains the age at which the policy expires. If a claimant still is disabled after he or she reaches the end of the benefit period, the waiver of premium provision will keep the policy in-force. The policy is still of value because he or she later may recover, return to work, resume payment of premiums, and then go on claim again. The benefit period is a limit on the amount of time the insured can collect for a single claim, not a limit on the total time the insured may collect during the life of the policy.\(^{12}\)

9 Modifications to the Model

With slight modifications, the model can be adapted to situations that do not fit all of the assumptions listed in Section 2.

\(^{12}\)During a claim audit, the author came across some insureds who twice had collected benefits successfully for the maximum length of time.
9.1 Nonannual Premium Payments

If premiums are paid \( j \) times per year, the units of insurance exposed to risk no longer will be constant between policy anniversaries. In this case, we need to make \( E_{\theta}(n) \) a function of the variable \( t \) as well as the variable \( n \). For the time being, let us ignore the effects of claim incidence and recovery. In this case, let \( q^{(w)}_{\theta}(n) \) be the policy withdrawal rate in the \( n \)th policy year and \( E_{\theta}^{(j)}(n, t) \) be the expected number of units in-force \( t \) (\( 0 \leq t \leq 1 \)) years after the start of the \( n \)th policy year given that premiums are paid \( j \) times per year. It easily is seen that

\[
E_{\theta}^{(j)}(n, t) = E_{\theta}^{(j)}(n, 0) \times \left(1 - \frac{[j \times t]}{j} q^{(w)}_{\theta}(n)\right),
\]

assuming lapses occur at the time of premium payment. The \([y]\) notation refers to the greatest integer less than or equal to \( y \).

9.2 Claim Incidence and Recovery

Insured lives constantly are migrating between the population exposed to risk for being on claim and the actual population on claim. In Section 2 and in equation (17) we chose to ignore this continual decrement and increment under the assumption that in any given policy year the population on claim is small relative to the active population. If this assumption is relaxed, however, the expected exposure is given by the following:

\[
E_{\theta}^{(j)}(n, \frac{k}{j} + t) = (1 - t r_{\theta}(n))E_{\theta}^{(j)}(n, \frac{k}{j}) \quad \text{for } 0 \leq t < 1/j \tag{18}
\]

\[
E_{\theta}^{(j)}(n, \frac{k + 1}{j}) = (1 - \frac{1}{j} (r_{\theta}(n) + q^{(w)}_{\theta}(n)))E_{\theta}^{(j)}(n, \frac{k}{j}) \tag{19}
\]

for \( k = 1, 2, \ldots, (j - 1) \). Again, we have assumed the uniform distribution of claims hypothesis.

Tracking the inflow of insurance units as claimants recover and return to work is more complex. In order to track this inflow continuously, we need an aggregate rate of claim recovery. Remember that the claim population consists of distinct cohorts of individuals who were disabled at different attained ages and at different policy durations. The rate at which a cohort recovers is a strong function of how long that cohort has been on claim. In general, the longer an insured has been on claim, the less likely he or she is to recover in the near future.
The aggregate inflow of recovering insureds is a mixture of lives from cohorts that are recovering at different rates.

To simplify matters, we will assume that premiums are paid once per year, i.e., \( j = 1 \). In addition, we assume that the net inflows and outflows over one year's time from policyholder mortality, claim incidence, and claim recovery are so small that we can wait until the end of the policy year to count them. So, instead of tracking the two way migration continuously, we only need to do so on policy anniversaries.

Define \( D_\theta(n, m) \) to be the number of units of insurance disabled in the \( n \)th policy year and still disabled by the beginning of the \( m \)th policy year.

**Case 1, \( m > n + 1 \):** An insured disabled between time \( n + t \) and \( n + t + dt \) must have been on claim at least \( m - n - e - t \) units of time in order to be disabled at the beginning of the \( m \)th policy year. Therefore:

\[
D_\theta(n, m) = r_\theta(n)E_\theta(n) \int_{t=0}^{1} s_\theta(m - n - e - t, n) dt. \tag{20}
\]

**Case 2, \( m = n + 1 \):** This case is different because some disabled insureds may not have completed the elimination period by the time the \( m \)th policy year has begun. In this case the value of \( D_\theta(n, m) \) is:

\[
D_\theta(n, m) = r_\theta(n)E_\theta(n)[e + \int_{t=0}^{1-e} s_\theta(1 - e - t, n) dt]. \tag{21}
\]

The first term in equation (21) accounts for those insureds who become disabled during the interval \((1 - e, 1)\) in the \( n \)th policy year. The net change in exposure due to incurrence/recovery is the sum of the change in the number of units on claim and the number of claims terminated in the preceding year by death.

### 9.3 Residual Disability

Up to this point we have assumed that a unit of insurance on claim is paid at the rate of $1 per year. If the insured is on residual\(^{13}\) disability, this may not be the case. Many disability income contracts sold today provide a residual disability benefit. Under such contracts, an insured can collect some fraction of his or her full benefit if a disability causes the individual to lose a portion of income but does not completely remove him or her from the work force. If an insured suffers a 60 percent loss of income because a disability renders him or her able to work only

---

\(^{13}\)In Goldman (1990) the term *residual* is used in a different sense. Group and individual terminology are not always equivalent.
part time or unable to do all of the duties of the occupation, he or she can collect 60 percent of the policy's full benefit. The benefit is a function of income lost, not amount of time unable to work. It is this fact that distinguishes the residual disability benefit from a partial disability benefit. A partial benefit is based on time lost rather than income lost. An insured who can put in a full day's work but loses income because he or she cannot perform certain key duties could qualify for a residual benefit.

To deal with residual disability benefits, we define a residual benefit function, $\rho_\theta(y, n)$ to be the fraction of the total benefits paid $y$ years from now to persons who were on residual disability in policy year $n$. For example, suppose 75 percent of all claimants begin their claims on total disability and the remaining 25 percent begin on 50 percent of the base benefit, then

$$\rho_\theta(0, n) = 0.75 \times 1 + 0.25 \times 0.5.$$  

Once $\rho_\theta(y, n)$ is known, we compute $BC_\theta(n, m)$ by multiplying $s_\theta(y, n)$ and $\rho_\theta(y, n)$ in the integrands. For example, in the case of residual disability, equation (14) becomes:

$$BC_\theta(n, m) = r_\theta(n)E_\theta(n) \int_{t=0}^{1} \int_{y=m-n-e-t}^{m+1-n-e-t} \rho_\theta(y, n)s_\theta(y, n)dydt.$$  

10 Analysis of Alternate Method

As mentioned earlier, Bluhm and Koppel (1988, Chapter 4, pp. 83–88) present a sample asset share calculation that purports to calculate benefit cash flows using claim costs and statutory claim reserve changes. We will point out deficiencies in this method and indicate how to correct the method. In addition, we will compare the Bluhm and Koppel method and the method presented in this paper.

Recall the notation used in Section 5.2. The equation used by Bluhm and Koppel can be restated as

$$BEN_n = CC_n - (CR^{(e)}_n - CR^{(b)}_n).$$  

(22)

There are two problems with this equation. First, if the $CR$ terms are the statutory reserve, then the equation is not based on assumptions that purport to be realistic. Statutory assumptions are supposed to be more conservative than realistic assumptions in order to ensure reserves contain a safety margin. For equation (22) to be correct, the $CR$
terms must denote the natural claim reserve, a reserve based on realistic assumptions.  

The second problem is that the change in claim reserve needs to be adjusted for interest. Equation (22) now will be corrected. Let \( R_\theta(n, m) \) be the natural claim reserve at the beginning of the \( m \)th policy year on insureds disabled in the \( n \)th policy year, let \( CC_\theta(m) \) be the claim cost incurred in the \( m \)th policy year, and let \( i \) be a realistic rate of interest. We will calculate \( BC_\theta(n, m) \), the benefit cash flow in the \( m \)th year to insureds disabled in the \( n \)th year.

**Case 1, \( n < m \):** Consider \( R_\theta(n, m) \) to be the current balance of a fund established to pay benefits to those insureds disabled in the \( n \)th policy year. Withdrawals are made to pay benefits. Interest is added, but no other deposits are made because the original balance of the fund exactly matches the present value of the benefits to be paid to the cohort of claimants. Then:

\[
R_\theta(n, m + 1) = (1 + i)R_\theta(n, m) - (1 + i)^{1/2}BC_\theta(n, m).
\]

We accumulate the reserve at the start of the year for a full year of interest and then subtract the money withdrawn to pay benefits (accounting also for the half year of interest lost when the withdrawal is made). The result is the fund balance at the end of the year. Rearranging this equation yields:

\[
BC_\theta(n, m) = -(1 + i)^{1/2}[vR_\theta(n, m + 1) - R_\theta(n, m)]. \quad (23)
\]

**Case 2, \( n = m \):** At the start of the \( m \)th policy year no one could have been disabled in the \( m \)th policy year. Thus, the fund balance on this empty cohort is zero. A deposit must be made to the fund when the cohort is established. That deposit is the claim cost. This gives

\[
R_\theta(m, m + 1) = (1 + i)^{1/2}[CC_\theta(m) - BC_\theta(m, m)]
\]

which yields

\[
BC_\theta(m, m) = -[v^{1/2}R_\theta(m, m + 1) - CC_\theta(m)]. \quad (24)
\]

---

This criticism does not strictly apply to the sample calculation presented by Bluhm and Koppel because in their example the claim reserves are calculated using pricing assumptions. For claims less than two years old, the valuation actuary is allowed to measure claim reserves using experience assumptions in place of the statutory valuation table. The Bluhm and Koppel example is for a disability income policy with a benefit period of two years. For benefit periods longer than two years, if the pricing actuary wishes to use this method, he or she will be obliged to calculate two sets of reserves: one realistic, the other statutory.
Summing the $BC_\theta(n, m)$ terms from $n = 1$ to $n = m$ yields the correct expression for $BEN_m$ (as opposed to the expression in equation (22)), i.e.,

$$BEN_m = CC_\theta(m) - (1 + i)^{1/2} \left( \nu CR_n^{(e)} - CR_n^{(b)} \right).$$  \hspace{1cm} (25)

Equation (25) may appear to be a lot easier to evaluate than those equations with double integrals in Section 7. This is not necessarily the case. To use the method in Section 7, the pricing actuary must calculate a two-dimensional array $BC_\theta(n, m)$. To project statutory book profit the pricing actuary also must project statutory claim reserves, among other things. To use equation (25), the pricing actuary must calculate a two-dimensional array $R_\theta(n, m)$ of natural claim reserves as well as an array of claim costs. If statutory claim reserves are different from natural reserves, the pricing actuary must project statutory claim reserves separately. The number of quantities to be calculated using equation (25) is larger, not smaller.

References


Appendix

The following example will demonstrate how the formulae can be evaluated to estimate cash flows. We project the amount paid in benefits in the first policy year and the amount of premium waived at the beginning of the second policy year. \( E_\theta(1) = 1000 \) units of insurance each with $100 per month of benefits ($100,000 per month in-force), \( e = 1/12 \) (30 day elimination period), \( w = 1/4 \) (90 day wait to qualify for waiver of premium), \( r_\theta(1) = 0.03 \), and tabulated values of \( s_\theta(y, 1) \) are given below:

\[
\begin{array}{cccc}
\text{Table A1} & \text{Data on } s_\theta \\
\hline
y & s_\theta(y, 1) & y & s_\theta(y, 1) \\
0/12 & 1.00 & 6/12 & 0.38 \\
1/12 & 0.80 & 7/12 & 0.36 \\
2/12 & 0.66 & 8/12 & 0.34 \\
3/12 & 0.54 & 9/12 & 0.33 \\
4/12 & 0.44 & 10/12 & 0.32 \\
5/12 & 0.40 & 11/12 & 0.31 \\
6/12 & 0.38 & 12/12 & 0.30 \\
\end{array}
\]

We can use equation (10) and a repeated trapezoidal rule (with monthly intervals) to calculate \( B C_\theta(1, 1) \).

\[
BC_\theta(1, 1) = \frac{1200}{144} r_\theta(1) E_\theta(1) \sum_{k=0}^{11-12e} \sum_{j=0}^{11-12e-k} \frac{1}{2} [s_\theta(j/12, 1) + s_\theta(j+1/12, 1)].
\]

From the data given, \( BC_\theta(1, 1) = 9,255 \).

Next we will estimate \( W_\theta(1, 2) \), the amount of premium waived at the start of the second policy year. Assume the annual premium rate, \( P_\theta \), is $10 per unit. From equation (15),

\[
W_\theta(1, 2) = P_\theta r_\theta(1) E_\theta(1) [w s_\theta(w - e, 1) + \frac{1}{12} \sum_{j=0}^{11-12w} \frac{1}{2} [s_\theta(1 - e - j/12, 1) + s_\theta(1 - e - j - 1/12, 1)]] = 132.00.
\]
Cost Containment in Workers’ Compensation: Evaluating Medical Fee Schedules

David L. Durbin and Barry I. Llewellyn*

Abstract

Medical expenditures in workers’ compensation programs have been subjected to few cost containment strategies. As workers’ compensation costs have escalated, however, increasing attention is being given to the role of medical fee schedules in containing the prices of medical services. To this end, we develop a model for estimating the potential cost savings from implementing medical fee schedules. A market basket of medical services received by injured workers is constructed. This basket is used to estimate the parameters of the model. In addition, the basket is used to determine the impact of imposing a fee schedule linked to usual and customary charges or to the Medicare resource-based relative value schedule (RBRVS).

Key words and phrases: basket, current procedural terminology, resource-based relative value schedule (RBRVS)

*David L. Durbin, Ph.D., is vice president in charge of claims and benefit research at the National Council on Compensation Insurance. He also is responsible for evaluating all legislative reforms affecting workers' compensation medical costs. Dr. Durbin completed his B.A. from Union College, Schenectady, New York and his M.A. and Ph.D. in economics from the City University of New York.

Barry I. Llewellyn, A.C.A.S., M.A.A.A., is senior vice president and actuary at the National Council on Compensation Insurance. He directs the activities on NCCI's law evaluations department that include responsibility for claims-related research, the determination of system cost drivers, and the definition and cost analysis of proposed and enacted reform initiatives.

Both authors' address: National Council on Compensation Insurance, 5 Marine View Plaza, Hoboken NJ 07030, USA.

The authors thank Anthony DiDonato, Norm Chretien, Chris Ritter, the editor, and the three anonymous referees for their comments and suggestions.

81
1 Introduction

The continuing rise in medical expenditures has led to significant public policy interest in methods to contain costs. Approximately $900 billion was spent on health care in the United States in 1993, slightly more than 14 percent of the United States' gross domestic product (Burnier, Waldo, and McKusick, 1992). Total health care expenditures have grown more than 10 percent per year over the past decade. Growth rates in expenditures also have been significant for the various social insurance programs where medical care either is funded directly by government programs (e.g., Medicaid and Medicare) or mandated by public policy (e.g., workers' compensation and private passenger automobile insurance).

For example, the medical cost portion of workers' compensation insurance, with payments of more than $20 billion in 1992, is estimated to have increased one and one-half times faster than health expenditures generally (Nelson, 1992). The absence of formal cost containment programs in workers' compensation may have contributed to this increase.

As medical benefits under workers' compensation insurance are essentially unlimited first dollar coverage\(^1\) with virtually no restrictions on covered services, there have been few opportunities for cost sharing.\(^2\) Cost-sharing programs have been used widely to contain group and individual health insurance costs. Newhouse, Phelps, and Marquis (1980) and Jacobs (1991) have shown that consumers (patients) of health care are price conscious, i.e., they respond to economic incentives provided through the price of care.\(^3\) This suggests that there is a downward sloping demand for medical care.

As with every other form of medical care, there has been increased attention paid to workers' compensation medical costs. There have been two main avenues by which cost containment traditionally has been pursued. The first involves setting fee limits on the specific medical services provided. Twenty-eight states currently have formal medical fee schedules in place regulating the price per service performed by

\(^{1}\)Unlimited first dollar coverage refers to coverage with no copayments, deductibles, or policy limits.

\(^{2}\)Over the past year or two, a couple of states have removed the prohibition on cost sharing in limited circumstances (e.g., where medical care is provided by a managed care network and the claimant has reached the point of maximum medical improvement.)

\(^{3}\)Research from the Rand health experiment estimates that the price elasticity of demand for health (when consumers have less than a 25 percent co-pay) is -0.2; thus, for a 10 percent increase in price to the consumer, consumption will fall 2 percent. Increases in the co-insurance rate also are found to affect consumption.
medical providers for care rendered to injured workers (Telles, 1993). The second involves restricting the injured worker’s choice of physician. The premise is that employers, who are financially responsible for providing medical benefits to injured workers, will have incentives to seek the most cost efficient medical providers. Twenty-two states currently allow employer selection of medical provider, and 41 states have restrictions on switching providers once an initial selection has been made (Telles, 1993).

The effectiveness of fee schedules and choice of physician on workers’ compensation medical costs may vary. It is clear from the little available research that each individual state program must be evaluated on its own merit. (See Durbin (1993) for a review.) Savings from these cost containment strategies have been observed, but the extent of such savings appears to depend on the individual state’s circumstances. In some situations savings may not be achieved. But there is continued interest in fee schedules and choice of physician as policymakers, insurers, and consumers seek ways to contain cost increases. Over the past year or so, more than 20 states have considered either changing their existing fee schedules or implementing new medical fee schedules.

This paper presents the methodology used by the National Council on Compensation Insurance (NCCI) for evaluating the cost effectiveness of medical fee schedules in workers’ compensation insurance. The data are based on the data provided to the NCCI for the evaluation of almost 100 different fee schedule proposals between January 1993 to April 1994. Several models are developed for estimating the potential cost savings for different types of fee schedules including the relatively new strategy of linking fees in workers’ compensation to the Medicare resource-based relative value schedule. We also discuss the development of a basket of medical services received by injured workers. Several common scenarios are presented as case studies. We conclude with a discussion of how estimated medical provider cost savings derived from fee schedules ultimately may affect the overall workers’ compensation system’s costs.

4The National Council on Compensation Insurance (NCCI) is a nonprofit data gathering, research, and ratemaking organization. NCCI compiles statistics on workers’ compensation and provides advisory rates or information to be used for establishing rates in 32 states. It also serves as statistical advisor in about half of the remaining states. NCCI is responsible for estimating the impact of workers’ compensation reforms, including fee schedules, in some 40 jurisdictions.

5Medicare is an United States social security program that provides health insurance protection for almost all Americans age 65 and over.
2 Fee Schedules

There are three general methods for establishing medical provider (as distinct from hospital or clinic) fee schedules: (i) fees may be set at some percentile of the usual and customary reimbursement rate (UCR); (ii) fees may be targeted at some multiple of the resource-based relative value schedule (RBRVS); and (iii) fees may be targeted at some multiple of another benchmark schedule (e.g., Blue Cross and Blue Shield). Any of these schedules may use specific information on actual charged amounts or, similar to RBRVS, they may use conversion factors to translate information on the relative resource use into dollar amounts. Some states use a combination of methods.

There are a number of factors that influence the design and construction of any fee or reimbursement schedule. The major objective is to contain costs through managing the price per service. Secondary objectives include equity (all patients/claimants charged the same for similar service) and administrative efficiency. Eccleston, Grannemann, and Dunleavy (1993) provide a comprehensive review of the design of workers' compensation fee schedules and provide interstate comparisons concerning these issues. In addition to the principal consideration of cost savings, they identify the following major design issues: specific levels of reimbursement and the basis for determining relative payment for procedures; coverage of providers and services; and provisions for updating prices. The study of such issues, while important, is beyond the scope of this paper as these questions are basically public policy issues. This paper deals with the evaluation of various fee schedules after they are designed.

Perhaps the most influential innovation in medical fee schedules has been the introduction of the RBRVS for Medicare. Designed by researchers at Harvard University the conceptual underpinning is straightforward: prices per service are designed to reflect the real resources (including time, equipment, and physician training) needed to perform the service. These relative values are adjusted for geographical differences in the cost of living. Thus, in essence, RBRVS is designed to reallocate resources among providers while retaining, at least as a first step, revenue neutrality.\footnote{For a brief overview of RBRVS, see Hsiao et al. (1988) or (1990a). For more detailed information, see Hsiao et al. (1990b), or write to Professor William C. Hsiao, Harvard University, School of Public Health, 1350 Massachusetts Ave., Room 726, Cambridge MA 02138, USA.}

A multiplicative factor of the RBRVS often is used in workers' compensation. The rationale is that if Medicare is able to offer neces-
sary and quality care, then this provides information on the minimum achievable price. Two states have implemented RBRVS for their workers' compensation systems (Pennsylvania and Vermont), while several other states have or currently are contemplating such an approach.\textsuperscript{7}

3 Evaluation Issues

Evaluating the impact of the imposition of a physician fee schedule is conceptually a straightforward exercise. In general, all fee schedules, regardless of the type of schedule, specify a maximum allowable charge for each service rendered by a medical provider.\textsuperscript{8} Services typically are identified by the current procedural terminology (CPT) guidelines published by the American Medical Association. The list of CPTs is extensive. For example, there are individual codes for different office visits (initial examination, follow-up visits), different surgical procedures (arthroscopy, laminectomy), medical tests, and physical therapy. There are five broad categories (medicine, physical medicine, radiology, surgery, and pathology) containing hundreds of individual service or CPT codes.\textsuperscript{9}

3.1 Percent of Usual and Customary

To estimate the statewide impact of imposing a fee schedule limiting the maximum allowable charge for each service requires three pieces of information: (i) information on the fee schedule amount for each CPT code; (ii) information on the amount that would have been charged for this CPT code in the absence of the fee schedule; and (iii) information on the distribution of services or the distribution of expenditures. Let $\Omega$ be the set of all different medical services (CPT codes) relevant to treating injured workers; $W_i$ be the premedical fee schedule expenditure weight or proportion of costs of the $i$th procedure in a given state; $F_i$ be the fee schedule or maximum reimbursement amount for the $i$th service; and $U$ be a measure of the usual and customary reimbursement rate

\textsuperscript{7}It is interesting to note that the American Medical Association recently supported the Medicare RBRVS approach as an alternative to price restrictions considered in the current health care debates in the United States.

\textsuperscript{8}Medical providers typically are defined by state statute. They include physicians, chiropractors, physical therapists, osteopaths, nurses, etc.

\textsuperscript{9}There is actually a sixth category for anesthesia. Anesthesia often is billed per unit of time, however, which complicates data reporting and analysis of reimbursement levels.
(UCR) for the $i$th service. The savings ($S$) can be represented as:

$$S = \sum_{i \in \Omega} \frac{W_i(U_i - F_i)}{U_i}.$$  \hspace{1cm} (1)

Note that in this formulation, if a particular service is not identified by a CPT code or if no schedule is specified (commonly referred to as a BR or by report), then $F$ is assumed equal to $U$ and no saving is attributed.

Stated differently, the empirical estimation of the maximum potential savings is a weighted average calculation where the weights are the proportions of total costs spent on each individual procedure.

Alternatively, $S$ may be estimated as:

$$S = \frac{\sum_{i \in \Omega} N_i(U_i - F_i)}{\sum_{i \in \Omega} N_i U_i} \hspace{1cm} (2)$$

where $N_i$ = the number of times the $i$th procedure is performed in a given state. The approach of equation (2) amounts to taking the difference between the total costs charged to all injured workers for a service before the fee schedule and the expected total costs after imposing the schedule as a proportion of total costs prior to the fee schedule.

The computation of $S$ requires extensive data, which often do not exist. With the exception of a few states that collect data on all services performed on injured workers, information on the number or distribution of services is not readily available. Similarly, aggregate expenditure information is generally not available. In many circumstances, little or no state-specific data are available. The difficulty in getting the appropriate data arises from the nature of workers' compensation claims. Each claim may consist of multiple treatments lasting a long period of time (i.e., a claim is defined as the entire disability period). In contrast, group and individual health insurance define a claim to be a single visit to a medical provider.

In order to estimate the savings potential from imposing a fee schedule in instances where there is little information, NCCI uses the concept of a basket. The NCCI basket has been compiled from five separate

10 Implicit in equations (1) and (2) is the simplifying assumption that the demand, the supply, and the distribution of medical services will not be altered by the imposition of a price ceiling (the fee schedule). As will be discussed later, there is strong evidence that challenges the appropriateness of this assumption. Violating this assumption has important implications for estimating the impact on total provider costs of imposing price limits per service.

11 The NCCI basket consists of a relatively small group of services that commonly is used in treating injured workers. It is analogous to the basket used by economists to construct price indexes such as the Consumer Price Index (CPI). Baskets have the advantage of providing reasonable approximations to the actual distribution or consumption of services without the prohibitive expense of massive data collection.
data sources containing information from 26 states with workers' compensation fee schedules. The data are derived from both private and government data sources and relate to medical services rendered between 1989 and 1991. Over 800 different CPT codes are represented; the NCCI basket combines the individual frequency data from all five sources into a aggregate frequency distribution. The basket is effectively a weighted average frequency distribution that provides the number and proportion of each service or CPT.

As noted above, the actual saving from a new fee schedule is the weighted sum of savings per service, with the weights provided by the basket. Because, by definition, the fee schedule provides the maximum reimbursement rate, the remaining piece of information relates to the current level of charges. Once the current distribution of charges and the frequency distribution of services are known, it is easy to calculate the total amount of reimbursements. Substituting the fee schedule levels for the UCR, it is again straightforward to calculate the expected reimbursements under the fee schedule. Using the basket, the estimated savings, $S_i^{(b)}$, are given by

$$S_i^{(b)} = \frac{\sum_{i \in \Omega^{(b)}} N_i^{(b)} (U_i - F_i)}{\sum_{i \in \Omega^{(b)}} N_i^{(b)} U_i}$$

where $\Omega^{(b)}$ is the set of services (CPT codes) in the NCCI basket; and $N_i^{(b)} = \text{the number of times the } i\text{th procedure in the basket is performed.}$

For example, Tables A1 and A2 (in the appendix) show an abbreviated hypothetical example of these calculations. Table A1 shows savings calculations by CPT. For the illustrated CPT codes, the following is shown: the NCCI frequency distribution or basket; the fee schedule level; the percentile of the UCR distribution where the fee schedule falls; the average of the UCR distribution up to the fee schedule level; and the average of the UCR. Based on these statistics, estimates of costs under both the current and new fee schedules and the estimated savings are shown.

The estimate of the new cost ($NC_i^{(b)}$) is:

$$NC_i^{(b)} = \rho_i^{(b)} N_i^{(b)} AVG_i^{(b)} + (1 - \rho_i^{(b)}) N_i^{(b)} FS_i^{(b)}$$

for $i \in \Omega^{(b)}$.

---

12 Because the CPT codes change periodically, the CPTs from earlier years are adjusted to reflect the CPT schedule as of 1991.
13 Data for the UCR, if not available from state agencies, often can be obtained from private medical bill audit vendors.
14 Throughout this paper the superscript $(b)$ notation is used to denote quantities calculated using only those services that are in the NCCI basket.
where, for \( i \in \Omega(b) \), we have \( N_i(b) = \) frequency distribution of service \( i \); \( \rho_i(b) = \) proportion of UCR that is less than the fee schedule's charge for service \( i \); \( \text{AVG}_i(b) = \) average charge of all charges that are less than the fee schedule's charge for service \( i \); and \( F_s(b) = \) fee schedule amount for service \( i \).

Table A2 summarizes the individual CPT calculations that yield the estimate of the overall system savings. In the example shown, the fee schedule is calculated to save 18.65 percent on total provider costs, with a range from 13.58 percent for the general medicine group to 37.58 percent for pathology. These represent estimates of the maximum potential savings on provider costs based on the restrictive assumption that the distribution of services will remain fixed.

### 3.2 Impact on Overall System Costs

In order to translate the impact of any fee schedule savings on physician costs to total workers' compensation system savings, two other pieces of information are needed. First, provider costs are only one component of overall medical costs. There are medical costs associated with at least two other broad categories: inpatient hospitalization and pharmaceuticals. Second, workers' compensation costs generally are separated into medical and indemnity (weekly disability benefits) components. Thus, information is needed on the proportion of medical costs that are provider-related and on the split between indemnity and medical benefits.

Table A2 provides information on the proportion of medical costs that are physician-related and the medical/indemnity split for the hypothetical example. In this example, physician costs are 48.4 percent of medical costs and medical costs are 58.0 percent of total costs. Thus, the estimated 18.65 percent savings on physician costs translates into a maximum workers' compensation system savings of 5.2 percent (0.0524 = 0.1865 \times 0.4841 \times 0.580).

NCCI has compiled information on the distribution of medical costs. These costs are split into three general categories based on information from four sources\(^{15}\) and are shown in Table A3. Based on the averages from the available data sources, medical costs are distributed as follows: 48.4 percent are physician costs; 47.6 percent are hospital costs; and 4 percent are other costs (including pharmaceuticals).

---

\(^{15}\)Information on the sources and the data on the distribution of medical costs are available from the authors.
3.3 RBRVS Evaluation

Evaluation of the potential impact of imposing the RBRVS in a particular state (or geographical region) requires essentially the same methodology as discussed in Section 3.1. The only difference is that the RBRVS must be converted into a maximum reimbursement amount for each CPT. Only after this conversion can comparisons be made to the current level of reimbursements in the market. The determination of the RBRVS amount for a particular CPT code for a given locale within a state requires that the values of three RBRVS parameters be known. The three RBRVS parameters are:

1. The conversion factor (CF) for the CPT code. There are three different conversion factors used for all states in 1994: $35.158 for CPT codes designated as surgical; $33.718 for CPT codes designated as primary care services; and $32.905 for nonsurgical services.

2. The relative value units (RVU) for the CPT code. The RVU for a given CPT is the same in all states. There are three RVUs for each code: a work RVU (RVU_w); a practice expense RVU (RVU_p); and a malpractice RVU (RVU_m).

3. The geographic practice cost index (GPCI) values that are specific to regions within a state. There are three GPcis for each region: a work index (GPCI_w); a practice expense index (GPCI_p); and a malpractice index (GPCI_m). The GPCI for any region is used for all CPT codes and is used to transform the national RVUs into local RVUs.

Let R be the RBRVS maximum payment for a CPT in a specific region; then

\[
RBRVS = \left[ (RVU_w \times GPCI_w) + (RVU_p \times GPCI_p) + (RVU_m \times GPCI_m) \right] \times CF. \tag{5}
\]

Tables A4 and A5 show these calculations for selected CPTs in a hypothetical state with three regions. Once these values are calculated, it is straightforward to compare them to current costs to estimate potential cost savings.

3.4 Updates to Existing Fee Schedules

The same methodology as outlined above may be used to evaluate updates or changes to existing fee schedules. Data availability issues
are not as pressing, because information will be readily available on both the old and new fee schedules. Equation (3) can be modified by replacing the measure of the UCR with the old fee schedule reimbursement level. Savings \( S_2^{(b)} \) are estimated as a weighted average as before, i.e.,

\[
S_2^{(b)} = \sum_{i \in \Omega^{(b)}} N_i^{(b)} \frac{F_i^{(old)} - F_i^{(new)}}{\sum_{i \in \Omega^{(b)}} N_i^{(b)} F_i^{(old)}}
\]

where \( F_i^{(old)} \) and \( F_i^{(new)} \) are the old and new service \( i \) fee schedules, respectively.

4 Utilization Issues

Economic theory and empirical research suggest that the demand, supply, and distribution of services will not remain fixed after implementation of a fee schedule. Microeconomic theory predicts that the imposition of fee schedules (price ceilings) for medical services will result in an increased demand for services and a reduction in the supply of these services. To meet this pent-up demand, substitutes to traditional medicine will be sought. Where substitutes are not readily available, alternative markets may result. The market for medical services is unique because of information asymmetries. Health care consumers must rely heavily on medical providers to prescribe the services they require. Thus, to some extent, physicians and other providers control the demand and the supply of care.

Research in health economics supports the notion that physicians may have a target or required income and that they act as profit maximizers (Ligon, 1994; Pauly, 1986). As a consequence, the imposition of a fee cap per service may alter physician behavior. Specifically, physicians may encourage so-called bracket creep where, instead of charging for a limited office visit, they charge for a comprehensive visit. In addition, physicians may resort to unbundling or billing for each specific service rather than grouping the services together. Given the prohibition on injured worker cost sharing in workers' compensation, these practices will erode the potential savings of the fee schedule.

\(^{16}\) This is not to suggest that black markets for medical care necessarily will develop as a result of price ceilings; the availability of substitute goods in the form of other diagnoses and treatments most likely will be affected. The point is to illustrate that alternative markets will develop to satisfy the demand. In the context of other types of regulation of medical care (notably the prohibition of certain medical procedures or services) black markets do develop. Examples include cancer treatments or abortions.
The federal government explicitly recognized this phenomenon in its imposition of the resource-based relative value schedule used in Medicare. Essentially, the federal government, based on a review of the health care economics literature (Federal Register, November 1991), estimated that as much as 50 percent of the anticipated fee schedule savings may be eroded and that the volume of services may be increased to at least partially offset the effect of the price cap.

There are two relevant studies that are workers' compensation specific. One conducted by the California Workers' Compensation Institute (1992) found that costs rose as a result of the imposition of a fee schedule. A more recent study by Roberts and Zonia (1994) corroborates the literature cited by the federal government. They found that in response to fee schedules, health care providers "tend to provide more complex procedures in a shorter period of time and tend to exploit ambiguities allowed under the fee schedule."

Because the above empirical analysis makes no explicit adjustments for the likely behavioral changes that may erode the savings potential, it may be appropriate to adjust the maximum savings level. For illustrative purposes, if a 50 percent adjustment factor similar to the one used by the federal government in RBRVS evaluations is applied in the hypothetical example, the anticipated overall workers' compensation system savings becomes 2.6 percent.

5 Summary and Conclusions

This paper provides a framework for evaluating the impact of imposing medical fee schedules on provider services. The methodology is straightforward. The constraints are data availability constraints rather than methodological constraints. In some (rare) instances states have detailed medical service data available on use and prices per CPT per injured workers. (Calendar year or fiscal year data are the only way data are captured.) If either piece of information is not available, then the analyst must rely on reasonable proxies.

Three pieces of information are required: frequency or expenditure weights per CPT; proportion of medical costs covered by the schedule; and a measure of current market prices — either UCR if evaluating a new fee schedule or information regarding the old fee schedule if evaluating a change in the fee schedule. Any schedules using relative values and conversion factors must be translated into dollar terms.

The NCCI methodology relies on a basket of services rendered to injured workers and further splits workers' compensation medical costs
into some general categories. This information is compiled from available data sources that, while generally suitable, may not be appropriate in all situations. In addition, the analyst must rely to a certain extent on judgment of the likely utilization offsets that will occur. Available research provides a guideline to probable behavioral changes.

Fee schedule design issues are important. The methodology presented will permit analysts to evaluate alternative scenarios, thereby providing policymakers and other workers' compensation insurance interested parties with benchmark information.

References


91/008B. Sponsored by the Health Care Financing Administration: Baltimore, Md., 1990b. (NTIS No. PB91-172197/XAB).


Table 1
Fee Schedule Analysis
Group = Medicine

<table>
<thead>
<tr>
<th>CPT</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90782</td>
<td>5,514</td>
<td>11.82</td>
<td>36.4428</td>
<td>9.811</td>
<td>15.225</td>
<td>$83,952</td>
<td>$61,139</td>
<td>$22,813</td>
<td>27.17</td>
</tr>
<tr>
<td>90843</td>
<td>5,688</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>63.190</td>
<td>$359,425</td>
<td>$359,425</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>90844</td>
<td>57,995</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>85.362</td>
<td>$4,950,569</td>
<td>$4,950,569</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>90853</td>
<td>5,052</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>55.430</td>
<td>$280,032</td>
<td>$280,032</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>90900</td>
<td>5,598</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>55.713</td>
<td>$311,880</td>
<td>$311,880</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>90915</td>
<td>6,964</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>55.713</td>
<td>$387,983</td>
<td>$387,983</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>95860</td>
<td>11,717</td>
<td>142.97</td>
<td>65.8729</td>
<td>104.327</td>
<td>135.040</td>
<td>$1,582,259</td>
<td>$1,376,924</td>
<td>$205,335</td>
<td>12.98</td>
</tr>
<tr>
<td>95861</td>
<td>6,279</td>
<td>188.12</td>
<td>46.2437</td>
<td>155.320</td>
<td>221.465</td>
<td>$1,390,578</td>
<td>$1,085,967</td>
<td>$304,611</td>
<td>21.91</td>
</tr>
<tr>
<td>95900</td>
<td>39,666</td>
<td>55.90</td>
<td>36.7641</td>
<td>47.827</td>
<td>71.301</td>
<td>$2,828,222</td>
<td>$2,099,596</td>
<td>$728,626</td>
<td>25.76</td>
</tr>
<tr>
<td>95904</td>
<td>24,114</td>
<td>40.85</td>
<td>5.4220</td>
<td>34.655</td>
<td>71.301</td>
<td>$1,719,350</td>
<td>$976,957</td>
<td>$742,394</td>
<td>43.18</td>
</tr>
<tr>
<td>99025</td>
<td>17,749</td>
<td>30.10</td>
<td>53.8303</td>
<td>23.985</td>
<td>33.784</td>
<td>$599,631</td>
<td>$475,824</td>
<td>$123,807</td>
<td>20.65</td>
</tr>
<tr>
<td>99075</td>
<td>14,307</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>307.127</td>
<td>$4,394,062</td>
<td>$4,394,062</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>99201</td>
<td>7,603</td>
<td>35.47</td>
<td>37.5369</td>
<td>30.755</td>
<td>38.312</td>
<td>$291,283</td>
<td>$256,221</td>
<td>$35,062</td>
<td>12.04</td>
</tr>
<tr>
<td>99202</td>
<td>131,439</td>
<td>44.07</td>
<td>36.5284</td>
<td>38.177</td>
<td>47.751</td>
<td>$6,276,311</td>
<td>$5,509,560</td>
<td>$766,752</td>
<td>12.22</td>
</tr>
<tr>
<td>99203</td>
<td>81,344</td>
<td>62.35</td>
<td>56.4976</td>
<td>53.186</td>
<td>62.187</td>
<td>$5,058,542</td>
<td>$4,650,640</td>
<td>$407,902</td>
<td>8.06</td>
</tr>
<tr>
<td>99204</td>
<td>11,871</td>
<td>75.25</td>
<td>20.5213</td>
<td>65.017</td>
<td>88.839</td>
<td>$1,054,603</td>
<td>$868,365</td>
<td>$186,238</td>
<td>17.66</td>
</tr>
<tr>
<td>99205</td>
<td>93,463</td>
<td>133.30</td>
<td>75.9654</td>
<td>108.538</td>
<td>119.932</td>
<td>$11,209,216</td>
<td>$10,700,563</td>
<td>$508,654</td>
<td>4.54</td>
</tr>
<tr>
<td>99211</td>
<td>16,257</td>
<td>24.72</td>
<td>75.7813</td>
<td>20.230</td>
<td>22.210</td>
<td>$361,062</td>
<td>$346,563</td>
<td>$14,499</td>
<td>4.02</td>
</tr>
<tr>
<td>99212</td>
<td>165,229</td>
<td>33.32</td>
<td>64.6289</td>
<td>27.789</td>
<td>31.561</td>
<td>$5,214,807</td>
<td>$4,914,835</td>
<td>$299,971</td>
<td>5.75</td>
</tr>
<tr>
<td>99213</td>
<td>492,230</td>
<td>38.70</td>
<td>46.4705</td>
<td>33.596</td>
<td>40.328</td>
<td>$19,850,678</td>
<td>$17,881,771</td>
<td>$1,968,907</td>
<td>9.92</td>
</tr>
<tr>
<td>99214</td>
<td>478,593</td>
<td>50.52</td>
<td>22.7081</td>
<td>43.825</td>
<td>58.446</td>
<td>$27,972,064</td>
<td>$23,450,905</td>
<td>$4,521,160</td>
<td>16.16</td>
</tr>
<tr>
<td>NCCI</td>
<td>Frequency</td>
<td>Maximum Fee</td>
<td>Maximum Fee Percentile</td>
<td>UCRP Average up to the Maximum Fee Percentile</td>
<td>Overall UCRP Average</td>
<td>Estimated Old Costs</td>
<td>Estimated New Costs</td>
<td>Estimated Savings</td>
<td>Estimated Savings as a Percentage of Estimated Old Costs</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>99215</td>
<td>176,853</td>
<td>74.17</td>
<td>13.7014</td>
<td>64.763</td>
<td>93.514</td>
<td>$16,538,290</td>
<td>$12,889,237</td>
<td>$3,649,053</td>
<td>22.06</td>
</tr>
<tr>
<td>99223</td>
<td>15,727</td>
<td>133.30</td>
<td>55.8102</td>
<td>110.784</td>
<td>140.107</td>
<td>$2,203,470</td>
<td>$1,898,781</td>
<td>$304,689</td>
<td>13.83</td>
</tr>
<tr>
<td>99231</td>
<td>27,968</td>
<td>37.62</td>
<td>35.1822</td>
<td>32.709</td>
<td>44.672</td>
<td>$1,249,385</td>
<td>$1,003,833</td>
<td>$245,552</td>
<td>19.65</td>
</tr>
<tr>
<td>99232</td>
<td>21,322</td>
<td>48.37</td>
<td>24.2045</td>
<td>42.196</td>
<td>60.916</td>
<td>$1,298,857</td>
<td>$999,481</td>
<td>$299,376</td>
<td>23.05</td>
</tr>
<tr>
<td>99233</td>
<td>8,302</td>
<td>64.50</td>
<td>11.9282</td>
<td>55.509</td>
<td>89.852</td>
<td>$745,947</td>
<td>$526,575</td>
<td>$219,372</td>
<td>29.41</td>
</tr>
<tr>
<td>99238</td>
<td>8,348</td>
<td>59.12</td>
<td>47.3425</td>
<td>50.676</td>
<td>65.993</td>
<td>$550,907</td>
<td>$460,161</td>
<td>$90,746</td>
<td>16.47</td>
</tr>
<tr>
<td>99243</td>
<td>11,399</td>
<td>*</td>
<td>*</td>
<td>106.618</td>
<td></td>
<td>$1,215,335</td>
<td>$1,215,335</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>99244</td>
<td>37,195</td>
<td>*</td>
<td>*</td>
<td>141.801</td>
<td></td>
<td>$5,274,306</td>
<td>$5,274,306</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>99245</td>
<td>36,646</td>
<td>*</td>
<td>*</td>
<td>182.316</td>
<td></td>
<td>$6,681,159</td>
<td>$6,681,159</td>
<td>$0</td>
<td>0.00</td>
</tr>
<tr>
<td>99281</td>
<td>20,726</td>
<td>34.40</td>
<td>17.9612</td>
<td>28.177</td>
<td>47.169</td>
<td>$977,627</td>
<td>$689,810</td>
<td>$287,817</td>
<td>29.44</td>
</tr>
<tr>
<td>99282</td>
<td>56,669</td>
<td>46.22</td>
<td>20.7523</td>
<td>37.388</td>
<td>60.473</td>
<td>$3,426,957</td>
<td>$2,515,375</td>
<td>$911,583</td>
<td>26.60</td>
</tr>
<tr>
<td>99283</td>
<td>50,132</td>
<td>56.97</td>
<td>9.0695</td>
<td>47.194</td>
<td>90.710</td>
<td>$4,547,466</td>
<td>$2,811,570</td>
<td>$1,735,896</td>
<td>38.17</td>
</tr>
<tr>
<td>99284</td>
<td>11,197</td>
<td>69.87</td>
<td>4.2828</td>
<td>58.738</td>
<td>136.669</td>
<td>$1,530,288</td>
<td>$776,996</td>
<td>$753,292</td>
<td>49.23</td>
</tr>
</tbody>
</table>

Note: Column headings are as follows: (1) NCCI Frequency; (2) Maximum Fee; (3) Maximum Fee Percentile; (4) UCRP Average up to the Maximum Fee Percentile; (5) Overall UCRP Average; (6) Estimated Old Costs; (7) Estimated New Costs; (8) Estimated Savings; and (9) Estimated Savings as a Percentage of Estimated Old Costs.

Estimated New Costs = [(1) × (3) × (4)] + [(100% - (3)) × (1) × (2)].

* denotes a “missing” value.
### Table 2

**Workers' Compensation Fee Schedule Analysis:**

**Effect on Overall System Costs**

<table>
<thead>
<tr>
<th>Group</th>
<th>NCCI Frequency</th>
<th>Total Estimated Old Costs</th>
<th>Total Estimated New Costs</th>
<th>Total Estimated Savings</th>
<th>Total Estimated Savings (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine</td>
<td>2,161,156</td>
<td>$142,416,503</td>
<td>$123,082,397</td>
<td>$19,334,106</td>
<td>13.58%</td>
</tr>
<tr>
<td>Radiology</td>
<td>651,524</td>
<td>$88,641,538</td>
<td>$60,125,246</td>
<td>$28,516,292</td>
<td>32.17%</td>
</tr>
<tr>
<td>Surgery</td>
<td>282,239</td>
<td>$140,650,077</td>
<td>$116,490,967</td>
<td>$24,159,110</td>
<td>17.18%</td>
</tr>
<tr>
<td>Pathology</td>
<td>200,365</td>
<td>$4,679,678</td>
<td>$2,921,047</td>
<td>$1,758,631</td>
<td>37.58%</td>
</tr>
</tbody>
</table>

**Total**    
7,730,791     $488,028,655          $397,018,190          $91,010,465       18.65%

*Phys, Med. = Physical Medicine.*

Percent Savings of Physician Costs: (from Total) 18.65%

Physician Average Costs as a Percentage of Total Workers' Comp. Medical Costs (from Table 3 ) 48.41%

Percent Savings of Total Medical Costs (0.0903 = 0.1865 x 0.4841) 9.03%

Medical Costs as Percent of Total Workers' Compensation System Costs (hypothetical value) 58.00%

Percent Savings of Total System Costs (0.0524 = 0.0903 x 0.58 ) 5.24%
Table 3
Distribution of Workers' Compensation Medical Costs

<table>
<thead>
<tr>
<th>Source</th>
<th>Physicians</th>
<th>Hospitals</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCCI Special Medical Study</td>
<td>57.80%</td>
<td>39.30%</td>
<td>2.90%</td>
</tr>
<tr>
<td>NCCI Closed Claim Study</td>
<td>44.53%</td>
<td>48.77%</td>
<td>6.70%</td>
</tr>
<tr>
<td>HCFA1</td>
<td>46.01%</td>
<td>50.31%</td>
<td>3.68%</td>
</tr>
<tr>
<td>NCIC2</td>
<td>45.30%</td>
<td>52.00%</td>
<td>2.70%</td>
</tr>
<tr>
<td>Average</td>
<td>48.41%</td>
<td>47.59%</td>
<td>4.00%</td>
</tr>
</tbody>
</table>

1HCFA = Health Care Financing Administration; administrative costs have been deleted from the HCFA data.
2NCIC = North Carolina Industrial Commission

Table 4
Calculation of RBRVS Fee Schedule Amounts

<table>
<thead>
<tr>
<th>1994 Conversion Factors</th>
<th>1994 Geographic Practice Cost Index Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVA CFV</td>
<td>Area PE MP</td>
</tr>
<tr>
<td>P $33.718</td>
<td>R1 1.053 1.139 1.231</td>
</tr>
<tr>
<td>S $35.158</td>
<td>R2 0.947 0.912 0.716</td>
</tr>
<tr>
<td>N $32.905</td>
<td>R3 0.956 0.980 0.716</td>
</tr>
</tbody>
</table>

Note: Column headings are as follows: IVA = Indicator Value; CFV = Conversion Factor Value; PE = Practice Expense; and MP = Malpractice;
Under column (IVA), the notation is as follows: P = Primary Care Services; S = Surgical Procedures; and N = Other Nonsurgical Services.
Under column (Area): R1 = Region 1; R2 = Region 2; and R3 = Rest of the State.
Table 5
Calculation of RBRVS Fee Schedule Amounts

<table>
<thead>
<tr>
<th>CPT</th>
<th>Geographically Adjusted RVUs</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RVUs</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22102</td>
<td>8.88</td>
<td>4.43</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1 9.351</td>
<td>5.046</td>
<td>0.776</td>
<td>15.172</td>
<td>$533.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2 8.409</td>
<td>4.040</td>
<td>0.451</td>
<td>12.901</td>
<td>$453.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3 8.489</td>
<td>4.341</td>
<td>0.451</td>
<td>13.282</td>
<td>$466.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63064</td>
<td>23.49</td>
<td>24.1</td>
<td>4.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1 24.735</td>
<td>27.450</td>
<td>5.096</td>
<td>57.281</td>
<td>$2,013.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2 22.245</td>
<td>21.979</td>
<td>2.964</td>
<td>47.188</td>
<td>$1,659.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3 22.456</td>
<td>23.618</td>
<td>2.964</td>
<td>49.039</td>
<td>$1,724.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99202</td>
<td>0.76</td>
<td>0.46</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1 0.800</td>
<td>0.524</td>
<td>0.062</td>
<td>1.386</td>
<td>$46.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2 0.720</td>
<td>0.420</td>
<td>0.036</td>
<td>1.175</td>
<td>$39.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3 0.727</td>
<td>0.451</td>
<td>0.036</td>
<td>1.213</td>
<td>$40.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99215</td>
<td>1.53</td>
<td>0.77</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1 1.611</td>
<td>0.877</td>
<td>0.086</td>
<td>2.574</td>
<td>$86.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2 1.449</td>
<td>0.702</td>
<td>0.050</td>
<td>2.201</td>
<td>$74.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3 1.463</td>
<td>0.755</td>
<td>0.050</td>
<td>2.267</td>
<td>$76.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>1.14</td>
<td>0.77</td>
<td>0.09</td>
<td>R1</td>
<td>1.200</td>
<td>0.877</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R2</td>
<td>1.080</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R3</td>
<td>1.090</td>
<td>0.755</td>
</tr>
<tr>
<td>99252</td>
<td>N</td>
<td>0.36</td>
<td>1.31</td>
<td>0.09</td>
<td>R1</td>
<td>0.379</td>
<td>1.492</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R2</td>
<td>0.341</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R3</td>
<td>0.344</td>
<td>1.284</td>
</tr>
</tbody>
</table>

Notes: Column headings are as follows: (1) Indicator; (2) Work; (3) Practice Expense; (4) Malpractice; (5) Area; (6) Work; (7) Practice Expense; (8) Malpractice; (9) Total; (10) RBRVS Fee.

In column (5), R1 = Region 1; R2 = Region 2; and R3 = Rest of the State.
The Markov Chain Interest Rate Scenario Generator Revisited
Sarah L.M. Christiansen*

Abstract
This paper furthers the development of the Markov chain interest rate generator. Though the basic technique remains essentially unchanged, there are still many significant changes to the model. For example: (i) the long (key) rates are now are generated by a mean reversionary process; (ii) the number of shapes is increased from seven to 11; (iii) the limitation of changing by only two shape codes per year is removed; and (iv) the random walk matrix that determines the shapes is revised to be more realistic. An algorithm is developed to determine the shape code of the original yield curve, thus eliminating an input and assuring consistency. Flexibility in the choice of the key rate is introduced. Implications of the choice of the key rate are discussed.

Key words and phrases: curve shape, yield curve, key rate, shape codes

1 Introduction

The Markov chain interest rate generator (MCG)\(^1\) was introduced by Christiansen (1992) as a model of the underlying term structure of interest rates. The MCG interest rate generator produces scenarios of

---

*Sarah L.M. Christiansen, Ph.D., F.S.A. (1991), M.A.A.A., received her B.A. in mathematics from the University of California, Riverside (1966) and her M.A. (1968) and Ph.D. (1971) from the University of New Mexico. Before starting her actuarial career in 1985, she taught mathematics at Drake University, Des Moines Area Community College, and Grandview College. She is currently an assistant corporate actuary at the Principal Financial Group, vice-chair of the Society of Actuaries' committee responsible for the F-480 examination entitled Derivative Securities: Theory and Application, and vice-chair of the Education and Research Section Council.

Dr. Christiansen's address is: The Principal Financial Group, 711 High Street, Des Moines IA 50392-0001, USA.

The author is grateful to Warren Adams, David Brown, Don Sanning, the editor, and the anonymous reviewers for their helpful suggestions.

\(^1\)An interest rate generator is an algorithm or an equation that produces scenarios of future interest rates.
complete yield curves for up to 30 years into the future. As a result it can be used to produce scenarios of spot yield curves of interest rates. These interest rates then can be used for cash flow testing, for asset/liability studies, for pricing purposes, to compare different investment strategies, or to test pricing assumptions for some annuity products. The MCG most often is used with rates that already incorporate the spreads earned over Treasury (risk-free) rates, but can be used with Treasury rates.

The objective of this paper is to present improvements to Christiansen's (1992) model. But before discussing these improvements, the basics of the original model are reviewed and summarized briefly in Section 2. (We hope that this review makes the current paper self-contained.) These improvements (discussed in Section 3) have increased the model's flexibility, while improving the realism of the yield curves and their distribution. Section 4 shows how the revised model can be implemented. The revised MCG model also has several other desirable characteristics (discussed in Section 5) such as producing reasonable rates and spot and forward rates that are never negative. Section 6 contains the conclusions and the appendices follow.

2 Review of Original Model

In the original model, the sequence of interest rates for the kth scenario\(^2\) is generated for \(n\) years and for different maturities using a particular long rate (or key rate). Without loss of generality, the scenarios were generated for \(n = 30\) years with a 20-year long rate, and for maturities of 1/2, 1, 2, 3, 5, 7, 11, 15, and 20 years. The sequence of interest rates for the kth scenario is determined in six steps.

Step 1, Get Sample of Uniform Random Numbers: For scenario \(k\), we obtain a sample of 30 uniform random numbers between 0 and \((R_{\text{max}} - R_{\text{min}} - 0.03)\)

\[ U_{k,t} \sim U(0, (R_{\text{max}} - R_{\text{min}} - 0.03)) \]

where \(t = 1, 2, \ldots, 30\); \(U(a, b)\) is the uniform distribution on \((a, b)\); \(R_{\text{max}}\) = the overall maximum rate to be permitted by the generator; and \(R_{\text{min}}\) = the minimum rate to be permitted.

Step 2, Determine MCG Key Interest Rate: Let

\[ \hat{I}_{k,t} = \text{MCG 20-year interest rate at time } t \text{ for scenario } k; \]

\(^2\)One hundred different scenarios were run.
The MCG interest rate at time \( t \) for scenario \( k \) is defined to be

\[
I_{k,t} = \max\{(1 - \rho)I_{k,t-1}, \min[(1 + \rho)I_{k,t-1}, 0.015 + U_{k,t}]\}
\]  

for \( t = 1,2,\ldots,30 \). The parameters \( R_{\text{max}}, R_{\text{min}}, \) and \( \rho \) are determined by the user. Christiansen (1992, p. 127) uses the following values:

\[
\rho = 0.20, \quad R_{\text{max}} = 0.25, \quad R_{\text{min}} = 0.03.
\]

**Shape Codes:** The shape of the spot yield curve is determined by the assumed shape of the curve. Seven shapes are identified from historical spot curves, and a symmetric envelope of shapes is identified. The shapes are coded from 1 for a steep normal (upward sloping) yield curve to 7 for a steeply inverted yield curve (see Figure 1). The user determines the code that best represents the shape of the original curve. The shape codes for each scenario then are determined iteratively. The original yield curve is used for all scenarios for time 0.

**Step 3, Get Random Walk Matrix:** Get the original random walk matrix of transition probabilities of moving from one shape yield curve to another shape yield curve from Table 1.

**Step 4, Determine Shape Code:** The random walk matrix then is translated into a *look-up matrix* containing the shape code for the previous curve; see Table 2. The shape code for the current curve under scenario \( k \) is determined as follows: let \( B_k = (b_{k,1}, b_{k,2}, \ldots, b_{k,30}) \) be a \( 1 \times 30 \) vector of random integers between 1 and 10 inclusive and \( \alpha_{k,t} \) be the shape code for scenario \( k \) at time \( t \). In addition, let \( L \) be the look-up matrix, i.e., \( L \) is a \( 10 \times 7 \) matrix with the \( i \)th row of \( L \) corresponding to the \( i \)th row of Table 2. For example, from Table 2 two elements of \( L \) are \( L_{1,7} = 5 \) and \( L_{10,4} = 6 \). The shape code, \( \alpha_{k,t} \), is found from the look-up matrix at the intersection of the column headed by the shape code for time \( t - 1 \) and the row containing \( b_{k,t} \), that is,

\[
\alpha_{k,t} = L_{i,j} \text{ where } i = b_{k,t} \text{ and } j = \alpha_{k,t-1}.
\]  

\(^3\)In this paper \( \rho \) is used instead of \( \text{ADJRMAX} \), which is the original notation of Christiansen.

\(^4\)Throughout this paper, only vectors and matrices are denoted in bold uppercase letters. Their elements are written in corresponding nonbold lower case letters. Only scalars are written in nonbold lower case letters.
Figure 1
Original Curve Shapes

Legend: (1) = steep normal; (2) = early peak; (3) = oscillating starting up; (4) = level; (5) = oscillating starting down; (6) = early valley; and (7) = steeply inverted.

Step 5, Determine the Intermediate Yield Curve $\hat{h}_{k,t}$: Each shape code $\alpha_{k,t}$ is associated with a vector of shape factors corresponding to points of maturity on the yield curve. In particular, $\alpha_{k,t}$ generates the row number of the factor shape table.

Let $\Omega$ be the set of maturity points given by

$$\Omega = \{1/2, 1, 2, 3, 5, 7, 11, 15, \text{ and } 20 \text{ years}\}$$

and let $G_{k,t}$ be the $1 \times 9$ vector of shape factors generated by $\alpha_{k,t}$ for the maturities of $\Omega$, i.e.,

$$G_{k,t} = (g_{k,t}(1/2), g_{k,t}(1), \ldots, g_{k,t}(m), \ldots, g_{k,t}(20))$$

for $m \in \Omega$. The factors for 20-year rates are given in Table 3. The product of these factors and the 20-year rate is the intermediate yield curve (a vector of rates), $\hat{h}_{k,t}$, i.e.,

$$\hat{h}_{k,t} = \hat{h}_{k,20} G_{k,t}.$$
Thus, for a specific maturity $m \in \Omega$, we have the intermediate rate is given by

$$\hat{I}_{k,t}(m) = \hat{I}_{k,20} \times g_{k,t}(m).$$

(4)

Step 6, Determine the Final Yield Curve $I_{k,t}$: The yield curve of final interest rates, $I_{k,t}$, are determined after checking the following conditions: (i) for each maturity $m$, the rates do not vary from the previous years' rate by more than a preset fraction ($\delta(m)$) of the previous rate; and (ii) the rates are subject to the overall maximum and minimum of $R_{\text{max}}$ and $R_{\text{min}}$, respectively. For each maturity, the final interest rates are given by

$$I_{k,t}(m) = \max\{R_{\text{min}}, (1 - \delta(m))I_{k,t-1}(m), \min[I_{k,t}(m), (1 + \delta(m))I_{k,t-1}(m), R_{\text{max}}]\}$$

(5)

for $m \in \Omega$. (Note that in Christiansen's original model, $\delta(m)$ is a constant, i.e., independent of $m$.)

<table>
<thead>
<tr>
<th>Old Shape</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Table 2

**Look-up Matrix**

<table>
<thead>
<tr>
<th>Random Number</th>
<th>Previous shape 1</th>
<th>Previous shape 2</th>
<th>Previous shape 3</th>
<th>Previous shape 4</th>
<th>Previous shape 5</th>
<th>Previous shape 6</th>
<th>Previous shape 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>


### Table 3

**Factors for Shapes: 20-Year Rate**

<table>
<thead>
<tr>
<th>SC</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.61</td>
<td>0.67</td>
<td>0.76</td>
<td>0.83</td>
<td>0.85</td>
<td>0.87</td>
<td>0.90</td>
<td>0.92</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>0.90</td>
<td>1.05</td>
<td>1.12</td>
<td>1.09</td>
<td>1.07</td>
<td>1.05</td>
<td>1.03</td>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>1.00</td>
<td>1.05</td>
<td>1.00</td>
<td>0.98</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1.06</td>
<td>1.00</td>
<td>0.95</td>
<td>1.00</td>
<td>1.03</td>
<td>1.04</td>
<td>1.02</td>
<td>1.00</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
<td>1.10</td>
<td>0.95</td>
<td>0.88</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.97</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1.39</td>
<td>1.33</td>
<td>1.24</td>
<td>1.17</td>
<td>1.15</td>
<td>1.13</td>
<td>1.10</td>
<td>1.07</td>
<td>1.04</td>
<td>1</td>
</tr>
</tbody>
</table>


*Note: SC = Shape Code.*
3 Improvements to the Original Model

We now discuss the improvements that have been made to Christiansen's (1992) model since the original research was done in 1989-1990.

3.1 Determination of the Key Rate Scenarios

The assumption of a uniform distribution of the 20-year rates is replaced by an updated form of the log-normal mean-reverting process. This process is used for the composite generator in Christiansen (1992). Though the uniform distribution appears to be effective in keeping rates in bounds without clustering at the bounds, there is no evidence in favor of a uniform distribution of interest rates. Doll (1991) does provide support, however, for the use of a mean reversionary interest rate generator.

All mean-reverting interest rate generators require a mean rate for the rate being modeled. It is considered desirable to have this mean be independent of the model's user. Murphy (1990, chapter 9, page 77) implies that the best goal is the current rate. This choice ensures that the mean rate is independent of the user. It also ensures that key rates are not biased upward when rates appear to be low or downward if recent history has periods of high rates.

For the kth scenario, the correction factor\(^5\) at time \(t\), \(CF_{k,t}\), for the mean-reversionary process selected is a parameterization of the one used by Jetton (1990). It is given by:

\[
CF_{k,t} = \text{sign}(D) \times SF_{k,t} \times \min(0.5|D|, 0.015|D^3|)
\]

where

\[
\text{sign}(D) = \begin{cases} 
1 & \text{if } D > 0; \\
0 & \text{if } D = 0; \\
-1 & \text{if } D < 0,
\end{cases}
\]

\(D = \text{mean rate} - \text{previous rate}\), and \(SF_{k,t}\) is the strength factor\(^6\) at time \(t\). The goal rate and the previous rate must be expressed as percentages,

\(^5\)A correction factor is used to bring outliers partially back toward the mean and to prevent the repeated use of normal random variables with mean zero and variance one \((N(0,1))\) from becoming \(N(0,n)\), which leads to unreasonable interest rates.

\(^6\)The strength factor is a number from 0 to 1 which impacts how fast the correction factor pulls back toward the mean. If the strength factor is too high, there is insufficient volatility in the rates, while a strength factor of 0 eliminates the mean-reversionary process and leaves a pure log-normal interest rate generator.
i.e., 6 percent is entered as 6 and not as 0.06. The strength factor may be constant or vary with time. The result is used as the mean-reversionary correction factor in the key rate calculation. Thus, for the kth scenario, the current mean-reversionary interest rate, \( MRI_{k,t+1} \), is given by

\[
MRI_{k,t+1} = (MRI_{k,t} + CF_{k,t})e^{\sigma Z_{k,t}}
\]  

(6)

where \( MRI_{k,t} \) is the previous mean-reversionary interest rate, \( \sigma \) is the volatility assumption, and \( Z_{k,t} \) is a random number from the standard normal distribution.

3.2 The Number of Shapes

The original seven shape factors include both a steep normal curve and a steep inverted curve and use maturities of 1/2, 1, 2, 3, 5, 7, 11, 15, and 20 years. Originally, there were no codes or probabilities for normal or inverted curves that were not steep, as it was thought that these shapes would result from the annual limitation on the permitted annual change. It became desirable to have shape codes and factors for the more usual forms of the normal curves and inverted curves, however, thus the increase to nine shapes. Also, the maturities used to specify the yield curve were changed to 1/2, 1, 2, 3, 4, 5, 7, 10, 15, 20, and 30 years because Barra7 changed the points that it specifies on its curves. In addition, there was a desire for the 30-year rate.

As the current curve steepened in 1993, it became clear that the steep normal curve is not sufficiently steep. Concern arose when it was noticed that the shortest rates always increased in the first year in all of the random scenarios. It was discovered that the product of the shape factor for the (then) steepest curve with the lowest possible long rate was above the current short rate! Part of the cause of this problem is the increase in the length of the curve from a 20-year maturity to a 30-year maturity and using a 30-year key rate. Two possible solutions were considered: increasing the permitted percentage change in the long rate or increasing the number of shapes. Increasing the permitted change permitted almost no decreases in the shortest rate, so this appeared not to be a viable solution and was discarded.

The second possible solution, increasing the number of shapes, involved a study of what the steepest shape should be. A historical database of curves was consulted, and the curves were examined on a factor basis. As historically there are steeper curves than the steepest shape in the generator, two more very steep positively sloped curve

---

7Barra is an international financial consulting firm.
shapes have been added to the generator, bringing the total number of curve shapes to 11. Adding new curve shapes solved the problem, i.e., the short rates now are allowed to decrease in the first year. A decrease in short rates does not happen often, as it still requires a drop in the long rate along with a shape code for the following rate that does not exceed that of the original curve. The possible shapes are no longer perfectly symmetric, as it was neither necessary nor desirable to add more steeply inverted curves. Figure 2 displays these new curves.

3.3 Permitted Annual Changes

The annual permitted change in rates originally consisted of two parameters: the permitted annual change in the long (key) rate and the permitted annual change in any rate. The smaller parameter is the permitted annual change in the key rate, and the larger parameter is the
permitted annual change in any rate. The new version has a permitted annual change for each rate, which, like the new volatility assumption, is a non-increasing function of the maturity. The requirement for an appropriate volatility assumption for any choice of the key rate led to the new volatility assumption.

### 3.4 Overall Maximum and Minimum Rates

The overall minimum and maximum rates are specified as parameters. Due to the lower interest rates in 1992 and 1993, regulators expressed concern that the minimum interest rate was too high. The regulators proposed a new dynamic minimum rate for the New York seven interest rate scenarios.\(^8\) The revised Markov chain generator made minor modifications to the new dynamic minimum rate for these scenarios. For assets other than Treasuries, the minimum is the lesser of 3.5 percent or 50 percent of the original rate for that maturity. For Treasury curves the minimum is the minimum of 2.5 percent and 50 percent of the original rate. The maximum rate remains level.

### 3.5 Frequency of Shapes

When the historical database of shapes was examined for steepness and for the changes of shape, it was discovered that the yield curve shifted in 1981 from an inverted yield curve to a normal curve and back again on a quarterly basis. Thus, it is too restrictive to permit only an annual change in shape codes of two or less. Therefore, in addition to increasing the number of shapes, the random walk matrix is restructured so that it is possible to move from one shape to any other shape. Many investment officers feel that some of the shapes (early peak and early valley) occur very rarely. Because it is desired to be as realistic as possible, the random walk matrix currently reflects the fact that these shapes are more rare than inverted curves.

The sum of probabilities assigned to the inverted curve shapes (8, 9, 10, and 11) is approximately the probability of an inverted curve as determined by Becker (1992). Becker determines inversions by using

---

\(^8\)The New York seven interest rate scenarios are the scenarios specified in New York Regulation 126. They are the following parallel shift scenarios: (1) no change; (2) rates rise 0.5 percent per year for the next ten years and then remain constant; (3) rates rise 1 percent per year for five years, then fall 1 percent per year for the next five years, and remain constant thereafter; (4) rates rise 3 percent in the first year and then remain constant. Scenarios 5–7 are the opposite of 2–4 rates falling instead of rising and rising instead of falling.
the ratio of the shortest rate to the 10-year rate; he considers a curve to be inverted if this ratio is at least 1.05. Because Becker does not look at the shape of the entire curve, it is difficult to determine the reasonableness of each individual curve type based on his data.

An algorithm to determine the shape code (at time $t = 0$) for the current curve has been developed. The algorithm requires that all rates be expressed as multiples of the key rate and exploits the idea that the short-term rates are more important in determining the shape than the longer term rates. Details are given in Appendix B.

Salomon Brothers Treasury bond equivalent yield curves,\(^9\) in yield to maturity form, are converted as closely as possible to equivalent spot yield curves. (It is necessary to use one 29-year rate instead of the 30-year rate.) These yield curves are added to the data obtained from Barra in spot yield form. Each of these 338 curves\(^10\) is run through the algorithm to determine its shape code. Eleven of these curves are graphed (in factor form). Figure 3 illustrates real rather than theoretical shapes. Table 4 shows the frequency of curve types contained in the sample of 338 curves.

<table>
<thead>
<tr>
<th>Curve Types and Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 1 2 3 4 5 6 7 8 9 10 11</td>
</tr>
<tr>
<td>FQ 3 16 30 107 14 39 69 20 2 36 2</td>
</tr>
</tbody>
</table>

*Note: CS = Curve Shape, and FQ = Frequency.*

### 3.6 Key Rate

The key rate is the one for which the original set of rates is generated by the mean-reversionary log-normal process. There are situations where the model is being used and a particular rate other than the long rate may be most important. In such a situation, it is possible to use that particular rate as the key rate. This flexibility requires the imposition of a volatility assumption for each maturity, which impacts both the rate generation process at the key rate (where the volatility assumption is used directly) and the adjustment process (where the permitted deviation from last year's rate varies by the maturity or duration of the rate studied).

\(^9\)Obtained from a database of Salomon Brothers yield curves from February 1965 through March 1993.

\(^10\)One per month from the above database for 28 complete years from February 1965 through January of 1993 plus February and March 1993.
The choice of the key rate determines the rate that will not be biased up or down in the process of generating rates. The key rate sequence is, by construction, equally likely to show increasing rates as decreasing ones. Unless the original yield curve is essentially flat, however, the shape of the original yield curve will impact the other rates. For example, when starting with the 20-year rate as the key rate and a steeply normal yield curve, the shorter rates tend to be higher than the original shorter rate. In the spring/summer of 1992 the yield curve was steeply normal, and using the 20-year rate as the key rate the generator produces 20-year rates that are higher than the original 20-year rate about 50 percent of the time. The 7-year rates produced are higher than or equal to the original 7-year rate about 70 percent of the time, however; see Appendix C.

When using the generator to test pricing, the 7-year rate may be a critical rate. Because one does not want to bias the 7-year rates up-
wards, the 7-year rate is selected as the key rate. Each new shape factor is the old shape factor divided by that curve’s key rate factor. Thus, the shape factors at the new key rate are all one.

When the 7-year rate was used as the key rate, however, the standard deviations of the resulting rates no longer are decreasing with increasing maturity; see Appendix C. This occurs even though the permitted percent change from one year to the next decreases with increasing time and the original volatility assumption is changed to depend on the key rate, and is non-increasing. When the longest rate is the key rate the standard deviations of the resulting rates generally decrease with increasing time until maturity. It does not appear to be possible to have both a key rate that is not the longest rate and to have decreasing standard deviations as the maturities lengthen. This appears to make intuitive sense from Figure 4. If the variation due to shape is squeezed out of one rate, it then must pop up in the others. Thus, there is a trade-off depending on which property is deemed to be the most important.

It appears historically that the long-term rate has been the most stable. Murphy (1990, Chapter 17, page 181) implies that the longest maturities have the lowest volatility (over the same time interval). As it is desirable from many standpoints to have decreasing standard deviations with increasing maturity (duration), the longest rate being modeled usually will be the key rate. This does have the effect of biasing the shorter rates: they will tend to go up more than down if the original yield curve is normally sloped and down more than up in the inverted situation.

4 Using the Revised Markov Chain Generator

The steps to follow are essentially the same as those described in Section 2 for the original model. The major difference is that, instead of using uniform random numbers, we use a sequence of standard normal random numbers; the number of shape codes is now 11. Thus, the sequence of interest rates \( \{I_{k,t}\} \) at time \( t \) for the \( k \)th scenario again is determined in six steps.

**Step 1, Get Normal Random Sample:** For the \( k \)th scenario, obtain a sample of 30 standard normal random numbers \( Z_{k,t} \) for \( t = 1, 2, \ldots, 30 \). Using these standard normal random numbers, compute the se-

---

11In the Salomon database, the 30-year rate had the lowest standard deviation.
Figure 4
Curve Shapes (7-Year Key)

Legend: (1) steep normal; (2) = normal; (3) = early peak; (4) = oscillating starting up; (5) = level; (6) = oscillating starting down; (7) = early valley; (8) = inverted; and (9) = steeply inverted.

Step 2, Determine MCG Interest Rate: The MCG 30-year interest rate at time $t$ for scenario $k$, $\tilde{I}_{k,t}$, is defined to be

$$\tilde{I}_{k,t} = \max\{ (1 - \rho)\tilde{I}_{k,t-1}, \min[ (1 + \rho)\tilde{I}_{k,t-1}, MRI_{k,t}] \} \quad (7)$$

for $t = 1, 2, \ldots, 30$, where $\tilde{I}_{k,0}$ = the current 30-year interest rate.

Step 3, Get Random Walk Matrix: Get the new random walk matrix of transition probabilities from Table 5.
Table 5
Random Walk Probabilities* for Shape Codes

<table>
<thead>
<tr>
<th>Original</th>
<th>New Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.220</td>
</tr>
<tr>
<td>2</td>
<td>0.120</td>
</tr>
<tr>
<td>3</td>
<td>0.100</td>
</tr>
<tr>
<td>4</td>
<td>0.090</td>
</tr>
<tr>
<td>5</td>
<td>0.090</td>
</tr>
<tr>
<td>6</td>
<td>0.080</td>
</tr>
<tr>
<td>7</td>
<td>0.060</td>
</tr>
<tr>
<td>8</td>
<td>0.040</td>
</tr>
<tr>
<td>9</td>
<td>0.020</td>
</tr>
<tr>
<td>10</td>
<td>0.020</td>
</tr>
<tr>
<td>11</td>
<td>0.010</td>
</tr>
</tbody>
</table>

*The probabilities along the rows sum to one.
Step 4, Determine Shape Code: Let $B_k = (b_{k,1}, b_{k,2}, \ldots, b_{k,30})$ again denote a vector of uniform random numbers between 0 and 1, and let $l = \alpha_{k,t-1}$ be the shape code for scenario $k$ at time $t - 1$.

Select row $l$ from the random walk matrix found in Step 4. Let $J(l) = (j_1(l), j_2(l), \ldots, j_{11}(l))$ denote the vector of cumulative sums of the 11 elements of row $l$. (Note $j_1(l)$ = first element of row $l$ and $j_{11}(l) = 1$.) Given $b_{k,t}$, the new shape code $\alpha_{k,t}$ is given by

$$\alpha_{k,t} = \min\{\alpha : j_\alpha(l) \geq b_{k,t}\}. \quad (8)$$

The shape code for all scenarios at time $t = 0$ is determined by the algorithm given in Appendix B.

New Shape Factors and Set of Maturities: Here the shape factors associated with $\alpha_{k,t}$ for 30-year rates are given in Table A1 in Appendix A. The new set of maturities is now

$$\Omega = \{1/2, 1, 2, 3, 5, 7, 11, 15, 20 \text{ and } 30 \text{ years}\}.$$

Steps 5 and 6: Given the new shape factors and set of maturities, $\Omega$, follow Steps 5 and 6 in Section 2.

5 Spot, Forward, and Yield to Maturity Rates

5.1 Positive Spot Rates and Forward Rates

Given the way this model is constructed, it produces only positive spot rates\(^{12}\) because every rate on the yield curve is the product of a positive long-term rate and a positive factor. Tilley (1992) points out that positive spot rates imply positive yield to maturity (coupon) rates, but not necessarily positive forward rates.\(^{13}\) Are any of the forward rates produced by this generator negative?

The relationship between spot rates and forward rates is as follows: let $s_n$ denote the $n$-year spot rate and let $f_n$ be the 1-year forward rate between years $n$ and $n + 1$. It follows that

$$f_n = \frac{(1 + s_{n+1})^{n+1}}{(1 + s_n)^n} - 1. \quad (9)$$

\(^{12}\)The $n$-year spot rate is the yield of a zero coupon (pure discount) bond maturing in $n$ years.

\(^{13}\)The $m$-year forward rate beginning $n$ years from now is the implied rate of interest to lend money $n$ years from now to be repaid in a single payment of principal and interest $m + n$ years from now.
Clearly, for a flat yield curve, the forward rates are equal to the spot rates. When the yield curve has a positive slope between two maturities, the associated forward rate will be positive and greater than the spot rate associated with the shorter maturity, i.e., \( f_n > s_n \), thus creating the expectation of rising rates with a positively sloped yield curve. Forward rates are lower than spot rates only when the rate for the shorter maturity is greater than the rate for the longer maturity, i.e., only in a negatively sloped portion of a yield curve.

If some spot rates are unknown, then the forward rate is defined differently. For example, assume that the spot rates \( s_k \) and \( s_n \) are known (with \( k < n \)) but the intervening spot rates are unknown. Then the assumption is made that the forward rates between \( k \) and \( n \) are all equal, giving

\[
f_j = \left[ \frac{(1 + s_n)^n}{(1 + s_k)^k} \right]^{1/(n-k)} - 1 \tag{10}
\]

for \( j = k, k+1, \ldots, n-1 \).

It is important that interest rate generators do not produce negative forward rates. Testing the theoretical shapes associated with this Markov chain generator produced no negative forward rates. The test was conducted by setting up a spreadsheet that had the factors for the shapes of the curves. When a 30-year rate was input, the spreadsheet computed \((1 + s_n)^n\) and successive differences of \((1 + s_n)^n\), i.e., \((1 + s_{n+1})^{n+1} - (1 + s_n)^n\). From equation (??), it is clear that a negative forward rate implies that these successive difference must be negative. All successive differences tested, however, were positive, implying that all forward rates must be positive. The successive differences were smallest when the longest rate was at a minimum. All of the successive differences were positive, even with the 30-year rate as low as 2 percent. Because the curves that were considered to be problematic are those with negative slopes, the effect of any minimum was ignored for this test only.

Further tests were made by considering that the shapes could vary from the theoretical due to the adjustment process, and these tests still did not produce zero or negative forward rates. Using the same spreadsheet, a positively sloped yield curve was set up for time \( t \), at a long-term rate of 3.5 percent, which is the minimum rate for this generator (ignoring minimums at other points). The following period it was assumed that the long rate had dropped by the maximum allowed and the curve had tried to invert. The resulting curve did not have a perfect shape, and this curve was tested to see if the successive powers of \((1 + s_n)^n\) were increasing. They increased regardless of the level of the long-term rate (6 percent also was tried, and they increased faster),
and each of the shapes of the positive curve was tried for the initial curve. Thus, it appears that this generator will produce only positive forward rates.

5.2 Spot Rates and Yield to Maturity Rates

The relationship between spot rates and yield to maturity rates depends upon the pattern of cash flows. Let \((y_1, y_2, \ldots, y_n)\) be the annual coupon payments from an interest-only bond maturing in \(n\) years at par ($1) and let \((s_1, s_2, \ldots, s_n)\) denote the sequence of spot rates. It follows that \(y_1 = s_1\) and

\[
1 = \frac{1 + y_k}{(1 + s_k)^k} + \sum_{j=1}^{k-1} \frac{y_k}{(1 + s_j)^j} \tag{11}
\]

for \(k = 2, 3, \ldots, n\). Equation (11) is based on the fact that an interest-only bond due at time \(k\) paying \(y_k\) annually and returning the principal at maturity should be valued at par using the spot yield curve. If the spot rates are known, the coupons easily can be found by solving equation (11) for \(y_k\), which gives

\[
y_k = \frac{[1 - (1 + s_k)^{-k}]}{\sum_{j=1}^{k} (1 + s_j)^{-j}}. \tag{12}
\]

If, on the other hand, the coupons are known, then the spot rates can be found easily by solving equation (11) iteratively. Thus, assuming that \((s_1, s_2, \ldots, s_{k-1})\) are known, then

\[
s_k = \left[ \frac{(1 + y_k)}{1 - y_k \sum_{j=1}^{k-1} (1 + s_j)^{-j}} \right]^{1/k} - 1. \tag{13}
\]

These formulas can be modified for interest rates and cash flows \(m\) times per year.

When equation (11) with the usual assumptions\(^\text{14}\) is applied to the historical yield to maturity curves for curve of December 1981, it yields a complex number for the value of the 30-year spot rate \((s_{30})\). (Note that Tilley (1992) mentions this problem when generators initially create YTM curves.) Because that result does not make sense, the 29-year rate is substituted.

\(^\text{14}\)The usual assumptions are (i) constant forward rates for intervals between specified points; (ii) semi-annual coupons; and (iii) compounding (adjusted for semi-annual payments and rates).
Using equation (??) is the only method available to convert curves when only one curve is available at a time. But the financial services and Wall Street firms do not determine the spot rate curves from the yield to maturity curves. Instead they determine the spot rate curves from the prices of on-the-run Treasuries, which are those priced closest to par. Their methods of fitting the spot rate curves are proprietary.

6 Conclusions

As with any other model, the Markov chain generator model continues to evolve as problems surface and questions arise. In order to further refine this model, a study is being made to reduce the number of scenarios from 1,000 to 50, while preserving the characteristics of the original set of scenarios. This study will test the results of the 50 scenarios and compare them to the original scenarios in cash flow testing.

References


Table A1
Factors for Shapes Using 30-Year Rate Key

<table>
<thead>
<tr>
<th>Code</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>0.42</td>
<td>0.51</td>
<td>0.57</td>
<td>0.64</td>
<td>0.71</td>
<td>0.76</td>
<td>0.80</td>
<td>0.83</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.48</td>
<td>0.53</td>
<td>0.61</td>
<td>0.68</td>
<td>0.73</td>
<td>0.77</td>
<td>0.82</td>
<td>0.85</td>
<td>0.88</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>0.59</td>
<td>0.64</td>
<td>0.73</td>
<td>0.80</td>
<td>0.82</td>
<td>0.84</td>
<td>0.87</td>
<td>0.88</td>
<td>0.92</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>0.74</td>
<td>0.78</td>
<td>0.83</td>
<td>0.85</td>
<td>0.89</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>0.78</td>
<td>0.94</td>
<td>1.09</td>
<td>1.17</td>
<td>1.14</td>
<td>1.11</td>
<td>1.09</td>
<td>1.07</td>
<td>1.06</td>
<td>1.04</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>0.96</td>
<td>1.02</td>
<td>1.07</td>
<td>1.02</td>
<td>0.99</td>
<td>0.98</td>
<td>1.00</td>
<td>1.02</td>
<td>1.04</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>1.04</td>
<td>0.98</td>
<td>0.93</td>
<td>0.98</td>
<td>1.00</td>
<td>1.02</td>
<td>1.00</td>
<td>0.98</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>1.20</td>
<td>1.06</td>
<td>0.91</td>
<td>0.85</td>
<td>0.88</td>
<td>0.89</td>
<td>0.91</td>
<td>0.93</td>
<td>0.94</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>1.28</td>
<td>1.22</td>
<td>1.17</td>
<td>1.15</td>
<td>1.13</td>
<td>1.09</td>
<td>1.07</td>
<td>1.05</td>
<td>1.04</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>1.51</td>
<td>1.47</td>
<td>1.39</td>
<td>1.34</td>
<td>1.30</td>
<td>1.28</td>
<td>1.25</td>
<td>1.23</td>
<td>1.16</td>
<td>1.09</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Table A2

**New Adjustment and Volatility Assumptions**

<table>
<thead>
<tr>
<th>Maturity</th>
<th>0.5 yr.</th>
<th>1 yr.</th>
<th>2 yr.</th>
<th>3 yr.</th>
<th>4 yr.</th>
<th>5 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Adjust.</td>
<td>0.400</td>
<td>0.350</td>
<td>0.330</td>
<td>0.300</td>
<td>0.270</td>
<td>0.250</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.170</td>
<td>0.165</td>
<td>0.160</td>
<td>0.140</td>
<td>0.130</td>
<td>0.120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturity</th>
<th>5 yr.</th>
<th>7 yr.</th>
<th>10 yr.</th>
<th>15 yr.</th>
<th>20 yr.</th>
<th>30 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Adjust.</td>
<td>0.250</td>
<td>0.200</td>
<td>0.160</td>
<td>0.130</td>
<td>0.100</td>
<td>0.100</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.120</td>
<td>0.115</td>
<td>0.110</td>
<td>0.100</td>
<td>0.100</td>
<td>0.095</td>
</tr>
</tbody>
</table>

### Appendix B: Algorithm to Determine Curve Shape

Let $\alpha$ be a number denoting the shape of the curve $C$, where $C$ is the $1 \times 11$ vector of the original rates. Also, let $M$ be an $11 \times 11$ matrix of shape factors and $M_i$ be the vector corresponds to curve $i$ (the $i$th row of $M$) for $i = 1, 2, 3, \ldots, 11$. The following is the algorithm to determine the shape of the original curve:

**Step 1:**  Let $R = C / s_{30}$ where $s_{30}$ is the 30-year rate.

**Step 2:** Make an element by element comparison of the vectors $R$ and $M_{11}$. If every element of $R$ is greater the corresponding element $M_{11}$, then set $\alpha = 11$ and go to Step 7. Otherwise, go to Step 3.

**Step 3:** Define the $1 \times 11$ vector $P$ and the $11 \times 11$ matrices $Q$, $V$, and $W$ follows:

$P = (0.1, 0.075, 0.05, 0.04, 0.035, 0.03, 0.02, 0.01, 0.005, 0.0025, 0.0)$;

$Q$ is a matrix where each row is $P$;

$V = M - Q$ and $W = M + Q$.

**Step 4:** Define a new $11 \times 11$ matrix $A$ as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } \nu_{ij} \leq r_j \leq \omega_{ij} \\ 0 & \text{otherwise} \end{cases}$$

for $i, j = 1, 2, \ldots, 11$.

**Step 5:** Define the $11 \times 1$ column vectors $D$ and $E$ as follows:

$d_i = (12 - i)^3$, $i = 1, 2, \ldots, 11$ and $E$ is the product $E = AD$. 
Step 6: \( \alpha = k \) where \( k = \max\{i : e_i \geq e_j, \text{ for } j = 1, 2, \ldots, 11\} \), i.e., \( k \) is the index with the largest value of the elements of \( E \).

Step 7: The shape code has been determined. Quit.

Appendix C: Summary Statistics

The following are summary statistics for all scenarios combined. The projection period is 30 years and all statistics are based on the entire period.

<table>
<thead>
<tr>
<th></th>
<th>7-Year Key Rate</th>
<th>20-Year Key Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-499 500-999</td>
<td>0-499 500-999</td>
</tr>
<tr>
<td>Mean</td>
<td>3.698 3.677</td>
<td>3.698 3.677</td>
</tr>
<tr>
<td>Median</td>
<td>2.000 2.000</td>
<td>2.000 2.000</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.686 2.661</td>
<td>2.686 2.661</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.000 1.000</td>
<td>1.000 1.000</td>
</tr>
<tr>
<td>Maximum</td>
<td>9.000 9.000</td>
<td>9.000 9.000</td>
</tr>
</tbody>
</table>

Note: 0-499 = Scenarios 0-499; and 500-999 = Scenarios 500-999.

Note that:

- For the scenarios 0-499 and the 7-year key rate, the percent of 7-year rates that is greater than the original rate is 49.22 percent, while the percentage equal to the original rate is 3.23 percent;

- For the scenarios 500-999 and the 7-year key rate, the percent of 7-year rates that is greater than the original rate is 46.74 percent, while the percentage equal to the original rate is 3.23 percent;

- For the scenarios 0-499 and the 20-year key rate, the percent of 7-year rates that is greater than the original rate is 68.55 percent, while the percentage equal to the original rate is 3.23 percent; and

- For the scenarios 500-999 and the 20-year key rate, the percent of 7-year rates that is greater than the original rate is 66.46 percent, while the percentage equal to the original rate is 3.23 percent.
# Key Rate: 7-Year Rate

## Table A4

### Scenarios 0–499: Statistics for Rates

<table>
<thead>
<tr>
<th>Rate</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mo.</td>
<td>7.4702</td>
<td>6.9429</td>
<td>2.8535</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>1 yr.</td>
<td>7.6977</td>
<td>7.2715</td>
<td>2.7584</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>2 yr.</td>
<td>7.9792</td>
<td>7.5970</td>
<td>2.6741</td>
<td>3.4161</td>
<td>25.0000</td>
</tr>
<tr>
<td>3 yr.</td>
<td>8.1034</td>
<td>7.7652</td>
<td>2.6267</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>5 yr.</td>
<td>8.3570</td>
<td>8.0155</td>
<td>2.6622</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>7 yr.</td>
<td>8.4634</td>
<td>8.3282</td>
<td>2.6841</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>11 yr.</td>
<td>8.5630</td>
<td>8.4110</td>
<td>2.6947</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>15 yr.</td>
<td>8.6241</td>
<td>8.4625</td>
<td>2.6825</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>20 yr.</td>
<td>8.7408</td>
<td>8.6612</td>
<td>2.6428</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
</tbody>
</table>

## Table A5

### Scenarios 500–999: Statistics for Rates

<table>
<thead>
<tr>
<th>Rate</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mo.</td>
<td>7.3108</td>
<td>6.7879</td>
<td>2.7803</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>1 yr.</td>
<td>7.5352</td>
<td>7.1258</td>
<td>2.6902</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>2 yr.</td>
<td>7.8172</td>
<td>7.4594</td>
<td>2.6187</td>
<td>3.4161</td>
<td>25.0000</td>
</tr>
<tr>
<td>3 yr.</td>
<td>7.9498</td>
<td>7.6245</td>
<td>2.5867</td>
<td>3.5000</td>
<td>24.2782</td>
</tr>
<tr>
<td>5 yr.</td>
<td>8.2021</td>
<td>7.9485</td>
<td>2.6299</td>
<td>3.5000</td>
<td>24.4737</td>
</tr>
<tr>
<td>7 yr.</td>
<td>8.3079</td>
<td>8.1680</td>
<td>2.6543</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>11 yr.</td>
<td>8.4083</td>
<td>8.2645</td>
<td>2.6647</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>15 yr.</td>
<td>8.4704</td>
<td>8.3297</td>
<td>2.6565</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
<tr>
<td>20 yr.</td>
<td>8.5976</td>
<td>8.5405</td>
<td>2.6280</td>
<td>3.5000</td>
<td>25.0000</td>
</tr>
</tbody>
</table>
Key Rate: 20-Year Rate

Table A6

<table>
<thead>
<tr>
<th>Rate</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mo.</td>
<td>8.2880</td>
<td>7.7566</td>
<td>2.8506</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>1 yr.</td>
<td>8.5445</td>
<td>8.0697</td>
<td>2.6599</td>
<td>2.9836</td>
<td>23.7613</td>
</tr>
<tr>
<td>2 yr.</td>
<td>8.8825</td>
<td>8.5735</td>
<td>2.4800</td>
<td>3.4161</td>
<td>23.5227</td>
</tr>
<tr>
<td>3 yr.</td>
<td>9.0386</td>
<td>8.7955</td>
<td>2.3625</td>
<td>3.5000</td>
<td>22.4026</td>
</tr>
<tr>
<td>5 yr.</td>
<td>9.3218</td>
<td>9.1213</td>
<td>2.3032</td>
<td>3.5000</td>
<td>22.1785</td>
</tr>
<tr>
<td>7 yr.</td>
<td>9.4352</td>
<td>9.2517</td>
<td>2.2665</td>
<td>3.5000</td>
<td>22.1785</td>
</tr>
<tr>
<td>11 yr.</td>
<td>9.5517</td>
<td>9.3750</td>
<td>2.2478</td>
<td>3.5000</td>
<td>22.4026</td>
</tr>
<tr>
<td>15 yr.</td>
<td>9.6199</td>
<td>9.4401</td>
<td>2.2387</td>
<td>3.5000</td>
<td>22.8506</td>
</tr>
<tr>
<td>20 yr.</td>
<td>9.7390</td>
<td>9.7140</td>
<td>2.2487</td>
<td>3.6004</td>
<td>22.4026</td>
</tr>
</tbody>
</table>

Table A7

<table>
<thead>
<tr>
<th>Rate</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mo.</td>
<td>8.1452</td>
<td>7.6545</td>
<td>2.7491</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>1 yr.</td>
<td>8.4010</td>
<td>8.0557</td>
<td>2.5637</td>
<td>2.9836</td>
<td>25.0000</td>
</tr>
<tr>
<td>2 yr.</td>
<td>8.7416</td>
<td>8.4416</td>
<td>2.3945</td>
<td>3.4161</td>
<td>24.7305</td>
</tr>
<tr>
<td>3 yr.</td>
<td>8.9069</td>
<td>8.6571</td>
<td>2.2933</td>
<td>3.5000</td>
<td>23.3344</td>
</tr>
<tr>
<td>5 yr.</td>
<td>9.1903</td>
<td>8.9933</td>
<td>2.2458</td>
<td>3.5000</td>
<td>22.5367</td>
</tr>
<tr>
<td>7 yr.</td>
<td>9.3034</td>
<td>9.1303</td>
<td>2.2124</td>
<td>3.5000</td>
<td>21.9435</td>
</tr>
<tr>
<td>15 yr.</td>
<td>9.4900</td>
<td>9.3163</td>
<td>2.1912</td>
<td>3.5000</td>
<td>22.6364</td>
</tr>
<tr>
<td>20 yr.</td>
<td>9.6128</td>
<td>9.6045</td>
<td>2.2095</td>
<td>3.5000</td>
<td>23.0984</td>
</tr>
</tbody>
</table>
Managing the Costs and Risks of Housing Finance: A New Role for Actuaries

Anthony Asher*

Abstract

Housing finance is a nontraditional field where actuarial expertise could be applied fruitfully. The development of mortgage instruments requires the application of financial mathematics, while the evaluation and management of the financial risks to which borrowers and lenders are exposed require a knowledge of insurance principles. This paper splits the financial costs of homeownership into several components: those that arise from inflation, risk, administration, and the residual real interest charge. The risk component further is partitioned into life contingencies, economic contingencies, and various moral hazards. This analysis provides a basis for future financial innovation, highlights where government intervention may prove productive, and suggests a number of areas of possible actuarial involvement.

Key words and phrases: mortgage, insurance, government guarantees, inflation, life cycle

1 Introduction

In their search for new areas of involvement, actuaries may find the field of housing finance both accessible and fruitful. The accessibility stems from the convergence of financial markets and the fading of the boundaries between insurers and mortgage finance institutions.

*Anthony Asher, F.I.A. (1980), is professor and director of actuarial studies at the University of the Witwatersrand, Johannesburg, South Africa. He previously was chief actuary of the Prudential Assurance Company of South Africa.

Professor Asher's address is: Department of Statistics and Actuarial Science, University of the Witwatersrand, Private Bag 3, 2050 WITS, Johannesburg, South Africa.

He thanks the anonymous referees and the editor for their comments and suggestions.
addition, there is a fundamental similarity between insurers and mortgage lenders: both are concerned with risk and security. Whether this becomes a fruitful area of expansion depends on the skills and energies of the actuaries concerned.

In any event, it would help actuaries if they better understood why some of the risks associated with the various types of home mortgages can, and even should, be insured. The need for mortgage insurance (for both lender and borrower) offers opportunities for those actuaries involved in product development. This paper is intended to highlight some of those areas in the development and evaluation of new mortgage products (for borrowers and lenders) where actuaries can play a role.¹

In the following sections, the basic types of mortgage products are examined. To facilitate this examination, the interest charged by mortgage lenders is divided into four components:

1. An allowance for inflation;
2. Charges for risk of default;
3. Costs of administration;
4. Net real interest.²

These components are discussed in Sections 2 to 5, respectively.

2 Inflation

2.1 The Need for Varying Payments

International norms are that housing loan payments should not exceed 25 percent to 30 percent of gross income and that the total amounts borrowed be no more than three to four years' annual salary. For a 20 year loan, these limitations are more or less compatible with nominal interest rates of 8 percent or less per annum. Once nominal interest rates exceed 8 percent (as is likely with double digit inflation rates), it is no longer possible to borrow three years' salary and still pay the

¹Financial innovation in home lending markets also includes variations on the theme of securitization (of mortgages) described by Woolford (1991) at a Society of Actuaries meeting. This subject has been covered extensively in the literature (see also, for instance, Schwartz and Torous (1992) and the March 1994 edition of Housing Finance International). It is not the business of this paper, however, which focuses more on the mortgage instruments themselves.

²The interest charge will include an allowance for the profit of the mortgage lender. The considerations involved in determining the profit are too complex to discuss here.
interest charges from current income.Extending the term of the loan makes little difference. A 50 percent increase in the term, for instance, will increase the amount that can be borrowed by only 14 percent (at an interest rate of 8 percent).

As the rate of inflation increases, it therefore would seem that lending institutions should be prepared to lend a portion of these interest payments to the borrowers. This would restore them to the cash flow position they had enjoyed before the rise in inflation, assuming that the present value of future income remains unchanged. Malpezzi (1990) gives examples from a number of countries of experiments with a variety of mortgage products that allow for greater flexibility and affordability, particularly in an inflationary environment.

One obvious solution is a mortgage instrument where a proportion of the nominal interest due is added to the amounts outstanding and the mortgage repayments are increased in nominal terms each year. Several of these instruments have been devised and used in different countries, with the rate of increase being determined in a variety of ways. Though there appears to be no standard nomenclature, these instruments seem to fit into three broad categories: low start mortgages, progressive annuity mortgages, and indexed mortgages. Formulae for the different types of mortgage instruments are given in the appendix.

Given their expertise in manipulating interest rates, cash flow testing, and matching assets and liabilities, actuaries should be able to make significant contributions in designing mortgage products and in matching mortgage assets with appropriate liabilities.

2.2 The Slow Response

Not all countries have institutions that have responded to an increase in nominal rates by introducing appropriate instruments. For example, the annual inflation rate has averaged over 14 percent for the past 20 years in South Africa, with nominal mortgage rates averaging just under 16 percent. Yet no increasing mortgage product has been publicly offered. Malpezzi (1990) shows the problem is not unique to South Africa.

Why have financial institutions been so slow in adapting? The answer probably includes elements of inertia, lack of expertise, and (quite proper) prudence. In his chapter that examines the effects of inflation on savings, Hadjimatheou (1987) records several studies that have found that higher inflation rates appear to create greater uncertainty and lead to greater savings by individuals. There is other evidence, presented later in this paper, that higher inflation may increase the risk of
borrowers defaulting on their loans. This lends some justification to the decisions by financial institutions to reduce their lending in times of higher inflation.

To understand the issues better, the question needs to be placed within the broader problem technically described by economists as *liquidity constraints within the life cycle*. Hadjimatheou quotes a summary of this idea by Hubbard and Judd (1986, p. 56):

> Hump-shaped lifetime earnings profiles rising towards middle age then leveling off and declining in old age imply that individuals will want to consume more than their current resources when young. They cannot do so if liquidity restraints are binding.3

Even in the United States, which has not experienced long periods of double digit interest rates, the literature provides evidence that higher nominal rates reduce the demand for debt.4 Megbolugbe and Linneman (1993) describe several studies that find a negative correlation between inflation and home ownership. These appear to be related to liquidity constraints. In economic terms, however, this represents a market failure. Begg, Fischer, and Dornbusch (1987) point out that market failure can occur as a result of the absence of appropriate insurance markets. They ascribe this absence to the presence of antiselection,5 moral hazard,6 and general lack of knowledge about the risks.

Actuaries have the expertise to address all of these questions (particularly those covering insurance) in a logically consistent manner.

---

3 Liquidity restraints result from unnecessary restrictions on the ability to borrow—in this case on the strength of future earnings.

4 This reduction in the demand for debt is consistent with a failure to introduce new mortgage instruments.

5 Antiselection is defined by the Society of Actuaries Committee on Actuarial Principles (1992) as a result of inadequate refinement of the risk classification system.

6 Moral hazard occurs when the future occurrence, timing, or severity of an insured event is controllable by the insured persons. Those risks that imply negligence rather than fraud are sometimes called morale hazard, but this distinction has not been used in this paper.
3 Risks

3.1 Some Background

3.1.1 Insurability: Private and Government

The conditions under which a risk (or event) is said to be insurable have been established by the Society of Actuaries Committee on Actuarial Principles (1992, p. 587). An event is insurable if:

1. It is associated with a phenomenon that is expected to display statistical regularity;

2. It is contingent with respect to number of occurrences, timing, or severity;

3. The fact of its occurrence is definitely determinable;

4. Its occurrence results in undesirable economic consequences for one or more persons; and

5. Its future occurrence, timing, or severity is neither precisely known nor controllable by these persons.

Moral hazards are not entirely unmanageable. Profit sharing, no-claim bonus, and the legal principle of indemnity all provide incentives to limit the number and size of claims. Careful definition of the insurable event that excludes those items under the control of the insured also allows for the extension of insurance markets. The courts can be used to challenge fraudulent claims.

Government attempts to insure, rather than to manage, the risks due to moral hazards are likely to result in unintended subsidies to unintended beneficiaries. Recent examples can be seen in the collapse of the savings and loans deposit insurance industry in the United States and the insolvency of the Motor Vehicle Accident Fund in South Africa. The latter has been driven to insolvency by, among other things, a large number of fraudulent claims.

The insurer can manage antiselection by obtaining sufficient information to refine its risk classification system adequately. Antiselection also can be managed through government intervention in the form of a mandatory insurance system. The need for personal information is reduced, if not eliminated.\(^7\)

\(^7\)Some persons' sense of equity might be offended if there were no differentiation between different risk classes. On the other hand, others might welcome the cross-subsidies involved.
Cyclical risks, such as unemployment, do not always display the statistical regularity required for insurability. Williams and Heins (1976) give an argument for government intervention to insure such risks. Where there is mandatory insurance, economic viability can be ensured by increasing premiums or reducing benefits.

Governments also may enter or create certain markets in countries where the private sector lacks the capital, technical capability, confidence, or energy to introduce economically viable financial products. There appears to be no inherent reason why such initiatives could not be privatized subsequently. Grigsby (1990) points out that in the United States mortgage guarantee insurance was introduced by government and later was followed by private initiatives.

3.1.2 Financial Planning

Appropriate responses to the risks faced by both mortgage lenders and borrowers must be developed. Duncan (1988) provides an interesting insight into the nature of the borrower’s risks. His paper is based on a detailed longitudinal study from 1968 to 1979 of 5,000 families and is called the “Panel Study of Income Dynamics.” He finds that family income is surprisingly volatile, with nearly one third of the sample experiencing a drop in income of 50 percent or more. (Family, in this context, can refer to individuals living on their own.) His analysis of the life events that cause some of these financial reverses will be used below. The results of the study highlight the need for financial planning and, where possible, appropriate insurance.

Bragg (1992) suggests that actuaries can find a new role in financial counselling. Are they not in the best position to know which risks are important, what insurance is available, and how finances should be managed?

Duncan’s paper and the life cycle hypothesis of Modigliani (1986) provide a good foundation for such planning. The issues, as mentioned earlier, are lifetime earnings and consumption and their variability, as well as current liquidity. The life cycle hypothesis assumes individuals behave rationally. But without assistance in planning their expenditures, it is difficult to understand how they can make rational decisions. Actuarial involvement in the design of lifetime income and expenditure models could provide a link between research and practical applications.

---

8It is the author's conviction that the individual's risks need to be considered first and that the need for the security of financial institutions derives from this primary concern.
Proper financial planning may indicate that the risk to the borrower of a home loan is larger than was previously thought and that the insurance necessary to manage the risk makes it too expensive. Thus, rather than making housing finance available to more persons, proper financial planning may reduce the demand for mortgages but fewer persons would suffer foreclosure.

3.1.3 Life and Disability Coverage

Mortgage lenders often require borrowers to obtain a credit death and disability life policy that covers the amount of the loan outstanding at the time of death or long-term disability. This may be inadequate, however, because death or disability can cause financial hardships of which meeting mortgage payments is only a part. The death and disability insurance policy rather should be designed with the family's future in mind. The family needs insurance to cover disaster and also may need assistance in rearranging its financial affairs after a claim. The institution's need to protect itself against loss probably can be met by a relatively small amount of life insurance (enough to cover the difference between the market value of the collateral and the outstanding loan if the latter is larger).

Duncan (1988) confirms that disability and illness are important causes of financial hardship. In his tabulation of the causes of the large declines in income, disability and illness are responsible in 12 percent of the cases for men between 25 and 45 and in 9 percent of the cases for women. This compares with only 3 percent for the death of a spouse, confirming the relative importance of disability compared to death. Unfortunately, disability coverage is not as readily available as life insurance. Lending institutions may be able to assist aspiring home owners obtain disability insurance by pooling borrowers into groups.

3.2 Cost Increases

Even if spared death or disability, borrowers still may find themselves unable to meet repayments because their costs and their incomes are not necessarily matched. This subsection looks at major reasons for cost increases, the next subsection examines other causes of income reduction.
3.2.1 Interest Rates

Holders of adjustable rate mortgages are particularly exposed to the risk of an increase in interest rates. Higher inflation rates appear to aggravate the problem; not only does the level of nominal interest rates rise, but short-term interest rates appear to display greater volatility (Chan et al. 1992).

Government protection from the risk of high interest rates, in the form of interest rate controls or subsidies, has not proved effective in volatile environments. Malpezzi (1990) argues against the use of such interventions. He finds that the distortions introduced into financial markets inhibit the appropriate allocations of resources without giving concomitant benefits to the poor.

Three mortgage products are commonly available as alternatives to government intervention: fixed interest mortgages; dual index mortgages; and privately subsidized mortgages.

- **Fixed Interest Mortgages**: Here the interest rate and the repayments remain unchanged throughout the term of the mortgage. This provides a level of certainty and probably makes this the most attractive mortgage product from the borrower's perspective. Such mortgages are particularly popular in the United States, but are regarded as too expensive in many other countries. Diamond and Lee (1992) show, however, that fixed interest mortgages are used in Denmark, France, and Germany (but without a prepayment option). This suggests that the product might be considered in other countries, at least those with relatively stable single digit interest rates.

Fixed interest mortgages, however, present a particular problem when it includes an option for borrowers to prepay the loan without penalty. Schwartz and Torous (1992) have calculated insurance fees for the prepayment risk, but it is not clear that the prepayments display sufficient statistical regularity to be insurable. Mortgage lenders at times have taken a large proportion of the risk themselves, but this can prove disastrous. Grigsby (1990)

---

9In an adjustable rate mortgage, interest rates and mortgage payments are fixed for a specified period (anywhere from one month to five years). At the end of this period, the outstanding balance is determined; new mortgage payments are recalculated using this outstanding balance and current market rates. The new rate and the new level of payments remain fixed for another specified period. This process is repeated until the mortgage is repaid. Adjustable rate mortgages are common outside the United States.

10Borrowers tend to repay their loans early when interest rates have fallen and they can refinance at a lower rate of interest.
Asher: Actuaries and Housing Finance

ascribes the American savings and loan debacle of the 1980s to, among other things, inadequate allowance for the mismatch of the term of their assets and liabilities caused partly by the pre-payment option.

- **Dual Index Mortgages**: Here the mortgage payments are increased in line with a wage-related index, while the cost of borrowing is related to market interest rates. Although not offering absolute protection against interest rate movements, the risk is deferred significantly and often will not give rise to losses. Dual indexed mortgages are more appropriate for countries with higher rates of interest; they are currently in favor with the World Bank. Roldan and Spoor (1992) tell how dual index mortgages have been introduced in Mexico with apparent success. A variant of the dual index mortgage was tried by one company in South Africa in the early 1980s: the lenders, however, suffered losses from loans still outstanding in the early 1990s. The problem was that the incomes of some borrowers did not keep pace with the required increases in the repayments. The loans of these borrowers became too large to be serviced and also exceeded the current market value of their homes.

- **Subsidized Mortgages**: The third approach is for employers to offer a fixed and subsidized lower rate of interest or to subsidize housing expenses on an income-related basis, as mentioned in Malpezzi (1990). This not only addresses the interest rate risk, but also the affordability of the loan. These subsidies, however, have become less popular over time as they are expensive and encourage overspending on housing.

Asher (1992) describes yet another form of mortgage: a salary-linked mortgage. This type of mortgage links both the periodic mortgage payment and the interest rate to the borrower's income. The borrower is liable to pay a predetermined proportion of his or her income over the term of the loan. The proportion is equal to the initial loan divided by the product of current income and the agreed term. The amount outstanding at any time is the product of the proportion calculated, the borrower's income at that time, and the remaining term. The interest earned by the lender in a salary linked mortgage is therefore dependent on the growth of the borrower's remuneration. The loan can be

---

11 In Mexico repayments are linked to the national minimum wage.

12 The concept of salary linkage is not that different from the income contingent loans described in Nerlove (1975) and elsewhere.
described as being linked to the index of the borrower’s own income. A portfolio of such loans is particularly appropriate for defined benefit pension funds as it provides an asset matched to the liabilities as both are dependent on income growth.

The link between income and mortgage payments introduces additional risks of antiselection and moral hazard. Antiselection may occur if individuals expecting lower salary growth are more likely to use such instruments. It should be feasible to develop classification models for potential borrowers to address this problem. Models can be developed that predict an individual's lifetime wage patterns.

Moral hazard risks are directly analogous to those that arise in collecting income tax. Income may be underreported (as in tax evasion), shifted to nontaxable sources (tax avoidance), and reduced by working less. Avoidance and evasion risks can be minimized by obtaining information on income from the employers who sponsor the pension fund making the loans.

Though the possible work disincentives cannot be gauged accurately, there are some studies of the problem. Tuomala (1990) reports that “most labor supply studies of men seem to indicate backward-sloping supply curves.” Higher income leads to men taking more leisure (described as an income effect), but the leisure is more expensive relative to other goods (which creates a substitution effect and reduces the leisure taken.) He lists 11 studies undertaken in the 1970s, of which seven showed the backward slope. Studies of women, however, usually have shown the normal forward slope. Brown (1983) gives more detail on some of these studies. A typical regression result with ten different parameters would produce an $R^2$ of 0.25, explaining only 25 percent of variation between individuals. Little confidence, therefore, can be placed in these results.

3.2.2 Individual Changes in Circumstance

Increased costs for the borrower can arise from many sources including divorce, inflation, illness, loss of or damage to possessions, increased family size, and the increased costs associated with children as they grow older. Duncan (1988) highlights the importance of divorce, where another household with its overhead costs is created. The problems may be aggravated by excessive borrowing—often credit purchases not considered when the initial housing loan was granted.

Higher levels of inflation appear to increase the possibility that cost increases may be unmanageable. Vining and Elwertowski (1976) confirm the widely accepted view that higher and more uncertain inflation
rates are linked with a wider dispersion of price movements. Their paper covers all elements of the Consumer Price Index, including housing. This wider dispersion may increase the chances of a family being unable to meet all of their financial obligations and justifies greater prudence when lending in times of higher inflation.

Mortgage lenders normally require borrowers to have homeowners' hazard insurance to protect their collateral against damage (from storms, fire, flooding, etc.). While such damage can lead to significant losses that otherwise would cause the borrower severe financial hardship, there are other risks to be considered such as unexpected medical costs to be considered. Thus, lenders may consider making health insurance a requirement in situations where adequate public care is not available.\textsuperscript{13} It is not possible, however, to insure against all fluctuations in expenditure. As a result, mortgage lenders still focus on the quality of the collateral rather than on the borrower's current and future income and expenses.

An actuarial model of lifetime income and expenditure that allows for variability in income and expenditure would be helpful in the evaluation of loan applications. Such a model also could be used to assist in rescheduling mortgage payments.

3.3 Reduction in Income

3.3.1 Partial Reduction

The risk of partial income loss\textsuperscript{14} appears to be greater with manual workers and older workers (those over the age of 50); see the surveys by Her Majesty's Stationery Office (1992) and Human Sciences Research Council (1990). Affordability standards imposed by lenders should allow for such fluctuations in income, particularly for older borrowers. At the present time in South Africa, a limited allowance sometimes is made for the risk of women's earnings declining as a result of childbirth, but not for other reasons. The salary-linked home finance product described above also addresses this risk directly.

\textsuperscript{13}Because individual health insurance policies are relatively expensive, mortgage lenders may consider pooling borrowers to provide them with access to group health coverage.

\textsuperscript{14}It appears that the problem of partial loss of income is the main reason for the difficulties experienced by the South African dual index mortgage scheme mentioned earlier. If wage patterns in Mexico follow those in South Africa, similar problems are likely for some Mexican borrowers in the second decade of operation of Mexico's dual index mortgage scheme.
The work of Lee (1988) in the United Kingdom shows that inflation increases the dispersion or real wage increases, but the earlier work of Hamermesh (1986) found the reverse to be true in the United States. The ambivalent findings of these authors indicate that the risks do not appear to have a regular statistical basis. This and the moral hazard involved in insuring items such as the number of overtime hours probably make the risk uninsurable.

3.3.2 Unemployment

Though governments offer some form of national unemployment insurance in several countries, there is little involvement by private insurance companies. This may be because unemployment appears to be subject to the three major causes of uninsurability referred to earlier: moral hazard, antiselection, and irregular statistical experience.\(^{15}\)

There is clear evidence of moral hazard. Schmitt and Wadsworth (1993), in reviewing the research on its presence in unemployment insurance, report that in the United States and the United Kingdom the duration elasticity (the percentage increase in average duration of unemployment related to a percentage increase in the unemployment benefit) appears to be some 0.4 and 0.3, respectively. It appears probable that the moral hazard would be even higher with an increase in the replacement ratio.\(^{16}\)

Antiselection is also probable. At any given time unemployment falls heavily on particular categories of workers (defined by geography, by industry, or by skills). These categories are themselves unstable; it would appear to be difficult for potential private insurers to predict these changes ahead of the workers concerned.

This degree of uninsurability provides a strong argument for some form of compulsory national unemployment insurance, which is desirable for reasons other than a reduction in the costs of housing finance.

Doling (1990) mentions that Finnish banks respond to short periods of unemployment by rescheduling repayments. He points out that this is possible because of the shorter terms (10 years) of the original loans. This is consistent with a life cycle planning model that takes future uncommitted income into account and with Duncan’s (1988) sample. In this sample, the major unemployment of the household head is a significant category, accounting for 19 percent of the large drops in income for men between 25 and 45. The importance of unemployment

\[^{15}\text{There is also a cyclical aspect to unemployment.}\]

\[^{16}\text{The replacement ratio is that of unemployment benefit to potential earnings.}\]
on defaults also is confirmed in studies by Canner et al. (1991) and Clauretice and Herzog (1990).

3.4 A Note on Mortgage Guarantee Insurance

Mortgage guarantee insurance protects mortgage lenders against losses on their lending. It normally is required for all loans issued by a lender, unless the borrower has made a sufficiently large down payment on the home so that the difference between the loan and the value of the home is large. Significant losses can occur in the event of a general fall in house prices that coincides with widespread income losses—due, perhaps, to a jump in the unemployment rate. The doubts expressed earlier (about the statistical regularity displayed by unemployment rates) must be magnified in the case of mortgage guarantees, as the movement of house prices appears even less statistically stable.

The insurance industry in the United Kingdom has suffered losses from mortgage guarantee insurance business of thousands of millions of pounds that may lead to a shortage of capacity in this market. These losses were particularly severe because housing prices in different regions of the United Kingdom tend to move roughly in tandem. Herzog (1988) shows that in the United States default risks (and presumably house prices) in the different regions have not moved together. Mortgage guarantee insurance, therefore, has been able to continue to provide security to mortgage lenders in the United States.

The author, however, has misgivings about mortgage guarantee insurance's long-term viability as an insurance product. If there is a prima facie case for its non-insurability, actuaries should consider the damage that a spate of claims could do to the insurance industry. They also could lobby for some compulsory coverage to be provided by government. As with unemployment, a compulsory scheme could adjust premiums and benefits in order to ensure solvency. A more thorough argument for government intervention is given by Foster and Herzog (1981), which could be a base for such lobbying. As Grossman (1992) points out, however, government guarantees introduce a moral hazard because the mortgage lenders no longer bear the cost of inadequate evaluation of borrowers.

17For example, in the United States borrowers must borrow no more than 80 percent of the appraised value of the home to avoid mortgage guarantee insurance.
3.5 Other Approaches to Managing the Default Risk

Statistical models of housing prices could be useful in determining appropriate collateral requirements for loans and reducing default probabilities. Gyourko and Voith (1992) and DiPasquale and Wheaton (1993) provide two recent attempts at developing such models.

Another approach would be to hedge the risks through financial options and futures on a house price index. Dwonczyk (1992) has developed just such an index in Australia. There might conceivably be persons hoping to buy homes in the future—in addition to financial speculators—who would find it attractive to take bull positions in these instruments. Although Dwonczyk expressed confidence that his index would be free of manipulation and accurately reflect the overall level of house prices, he also reported that adaptations were required for the 23 submarkets he found in Sydney. It is unlikely that a market could develop in derivative instruments for each of these submarkets, which would limit their usefulness as hedging instruments.

4 Administration

4.1 Initial

Most of the costs of lending are front ended, being expended in selling, evaluating the collateral, and performing the legal and administrative procedures necessary to initiate loans. Various stamp duties and transfer taxes also may be payable. The costs of administration normally are charged as a percentage of the amount outstanding.

It is, therefore, necessary to estimate the average term of the loan in order to decide on a reasonable amortization period for the initial costs. The term will be longer if the mortgages are transferable (when the borrower moves to another house) or assumable (when the mortgage is assumed by the new owner).

Charging the borrower directly for all the initial costs would be consistent with the approach taken in the life insurance industry. It has the advantage of eliminating cross subsidies to borrowers who prepay and shifts the risk of underestimating the likely term of the loan from the institution to the borrower.18 It is likely to lead to lower charges, but because it emphasizes the initial costs, it may repel some prospective customers.

18In the United States, lenders do reduce the impact of prepayment of mortgages by requiring that initial expenses be paid in advance.
Ameliorating the tilt\textsuperscript{19} problem in countries with relatively high rates of inflation also should reduce the annual charges for initial costs. This is because there is less need to enter the lower end of the housing market and to trade up as larger dwellings become affordable.

4.2 Renewal

The ongoing costs of installment collection are relatively small and particularly amenable to automation. Berger and Humphrey (1991) find that efficiencies in administration differ markedly from bank to bank in the United States, but are not especially related to size. The critical issue is limiting input costs. This appears to accord with the South African experience. Mergers and acquisitions appear not to be appropriate strategies for the control of expenses.

4.3 Final

Final expenses are fairly trivial if the loan is fully repaid on or before due time. Costs can be considerable in the event of defaults and foreclosure. Insurance of the risks leading to foreclosure will reduce these costs. The administration costs of foreclosure should be added to the losses incurred as a result of the failure of the collateral, as a saving in one can create costs in the other.

5 Real Interest Charges

The real interest charge represents the balance between demand for, and supply of, loanable funds. It accords with the risk-free interest rate of financial theory and normally is regarded as the rate available on short-term government bills.

Financial innovation by mortgage lenders is unlikely to yield any long-term advantage over other financial institutions in reducing the cost of the real interest rate—any innovations can be copied. The best that can be expected is for the lending institutions' financial instruments on the liability side of their balance sheets to be up-to-date so that housing interest rates are comparable with the rest of the market. The development of secondary markets may be a necessary part of this activity. Actuaries could have a role in this area.

\textsuperscript{19}Level nominal mortgage repayments are tilted once they are adjusted for inflation.
Governments can, and frequently have, intervene(d) in financial markets to reduce the costs of housing finance. Mayo (1991), in summarizing the World Bank's view on housing, suggests that governments should focus on enabling markets to function and that positive real interest rates are more appropriate. It is difficult to ensure that the subsidies inherent in artificially low interest rates help those they are intended to help.

6 Concluding Remarks

Numerous cost effective housing finance instruments have been designed to adapt to unpredictable inflation. In addition, many conventional insurance products are available to cover the risks to both borrower and lender. Proper financial planning would indicate that these insurance coverages are required by borrowers whether or not they have home loans. Regarding these risks as part of the business of the mortgage lender represents, in some respects, an artificial increase in the cost of housing finance and can have the undesirable effect of unnecessarily dispossessing the homeowner.

Actuaries could have two roles in managing the cost of housing. First, there is the need for better financial planning and for new insurance products. Second, actuaries could find new fields in the employ of home lenders, especially in financial management and product development.

References


Appendix

The present value of annual payments in arrears over \( n \) years is given by:

\[
\text{Value} = X \sum_{k=1}^{n} \prod_{i=1}^{k} \frac{(1 + f_i)}{(1 + g_i)}
\]

where \( X \) is the unadjusted base payment, \( X(1 + f_1) \) is the first payment, the interest rate in year \( i \) is \( g_i \), and the installment is increased by \( f_i \) in year \( i \).

Some of the more commonly used types of mortgages are defined in terms of the rules used to determine \( f_i, g_i, \) and \( n \) (\( n \) is assumed to be fixed unless otherwise stated).

- **Fixed interest mortgages:**
  
  \[
  f_i = 0 \quad \text{for all } i; \\
  g_i = g, \text{ a fixed rate of interest for all } i.
  \]

- **Adjustable rate mortgages:**
  
  \[
  f_i = \frac{a(n - i, g_{i-1})}{a(n - i, g_i)} - 1 \quad \text{for all } i; \\
  g_i = \text{The market rate of interest in year } i
  \]
  
  where \( a(n, r) \) is the present value of an \( n \)-year annuity immediate at a rate of interest of \( r \) per annum, i.e.,

  \[
  a(n, r) = \frac{(1 - (1 + r)^{-n})}{r}.
  \]

- **Low-start mortgages:**
  
  \[
  f_i = \text{Some low, predetermined, rate of increase for } (i < k); \\
  f_i = \frac{a(n - i, g_{i-1})}{a(n - i, g_i)} - 1 \quad \text{for } i \geq k; \\
  g_i = \text{The market rate of interest in year } i.
  \]
• **Progressive annuities:**

\[
\begin{align*}
  f_i &= \frac{(1 + g_i)}{(1 + j)} - 1 \quad \text{where } j \text{ represents the real rate of interest;} \\
  g_i &= \text{The nominal rate of interest in year } i.
\end{align*}
\]

• **Price level adjusted mortgages:**

\[
\begin{align*}
  f_i &= \text{The rate of inflation in year } i; \\
  g_i &= (1 + f_i)(1 + j) - 1 \quad \text{where } j = \text{the real interest rate.}
\end{align*}
\]

As a further generalization, these mortgages can be viewed as index-linked mortgages where \( f_i \) is linked to any index.

• **Dual index mortgages:**

\[
\begin{align*}
  f_i &= \text{Growth rate of index (usually related to wages) in year } i; \\
  g_i &= \text{Interest rate in year } i
\end{align*}
\]

As \( n \) is no longer fixed, the series may not have a finite value.

• **Salary linked home finance:**

\[
\begin{align*}
  f_i &= \text{The rate of growth of the borrower's salary} \\
  g_i &= f_i + r_i
\end{align*}
\]

where \( r_i \) is some nonnegative rate of interest in excess of salary growth.
Reconciling Two Rate Level Indications: A Chain Rule Approach

Cheng-Sheng Peter Wu*

Abstract

The problem considered is that of reconciling two rate level indications that are based on several common factors, but have been made at different review periods. A popular approach to this problem is the so-called sequential replacement method, which calculates the impact of each individual factor. Unfortunately, this method has a serious deficiency: the estimated impact of a factor depends upon the order of the replacement. To counteract this defect, a new approach, called the chain rule approach, is developed. Using this approach, an explicit formula is given for calculating the impact and the marginal impact of each factor.

Key words and phrases: sequential replacement approach, factor impact, marginal factor impact

1 Introduction

Pricing (property/casualty) actuaries often have to deal with situations where two rate level indications have been produced at different rate review periods. If the two indications differ, then underwriters, marketing personnel, and regulators usually want to know the reasons

*Cheng-Sheng Peter Wu, A.C.A.S. (1993), is a senior actuarial associate at Aetna Life and Casualty, Hartford, CT. He received his masters degrees in chemical engineering and in statistics from the Pennsylvania State University. Mr. Wu has several papers published (or in press) in the areas of automotive engineering, tribology (lubrication engineering) and statistics.

Mr. Wu's address is: Personal Auto SBU-TS24, Aetna Life & Casualty, 151 Farmington Ave., Hartford CT 06156, USA.

He thanks the anonymous referees and the editor for their detailed comments and suggestions.
behind the difference. This is particularly true when rate level indications increase 20 percent, 30 percent, or even 50 percent in one year for volatile lines such as workers’ compensation. Such significant increases may result from various factors, including a high trend, deteriorating experience, or a change in the loss development pattern.

In order to explain the difference between two rate level indications, the pricing actuary may need to estimate the individual impact of each rating factor on the change. Because the rate level indication function is usually a nonlinear function of the underlying rating factors, the portion of the overall change due to any given factor depends on the values of other factors.

In Section 2 we describe the approach now in use, the so-called sequential replacement approach. A better method, called the chain rule approach is introduced in Section 3.

2 The Sequential Replacement Approach

2.1 The Definition

One method that some actuaries use to reconcile two rate level indications is the sequential replacement approach. The sequential replacement approach starts with the prior review indication and replaces the prior review rating factors sequentially (one by one) with the current review factors. The method then concludes that the impact of any factor is the change in the indication when that particular factor is replaced in the indication calculation. In other words, suppose that there are \( m \) factors and at time \( t \), for \( t = 0, 1, 2, \ldots \), the \( i \)th factor\(^1\) is denoted by \( x_{t,i} \) and the vector of the \( m \) factors at time \( t \) is denoted by \( x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,m}) \). The indication at time \( t \) is \( I_t \), given by

\[
I_t = f(x_t)
\]

where \( f \) is a real valued function of the \( m \) factors. The change in the indication is \( \Delta I_t = I_{t+1} - I_t \). How do we calculate the change in indication due to a particular factor? According to the sequential replacement method, the change in indication between times \( t \) and \( t + 1 \) for factor \( i \) is \( \Delta I_t(i) \) where

\[
\Delta I_t(i) = f(x_{t+1,1}, x_{t+1,2}, \ldots, x_{t+1,i-1}, x_{t+1,i+1}, \ldots, x_{t+1,m}) - f(x_{t+1,1}, x_{t+1,2}, \ldots, x_{t+1,i-1}, x_{t,i}, x_{t,i+1}, \ldots, x_{t,m}).
\]

---

\(^1\)The factors can be labeled in any order, provided the order is maintained throughout the analysis.
An obvious problem with using equation (2) to measure the impact of a specific factor is that the size of the estimated impact depends on how the factors are labeled and the order in which they are replaced in equation (2). The following example will illustrate this problem.

2.2 Example 1: The Sequential Replacement Approach

Assume that the following generic loss ratio formula\(^2\) is used to calculate a rate indication:

\[
I = \frac{X \times C + (1 - C) \times B}{ELR} - 1
\]  

(3)

where \(I\) is the indication; \(X\) is the insurer's ultimate, on-level, and trended experience; \(C\) is the credibility; \(B\) is the experience applied to the complement of credibility; and \(ELR\) is the permissible or expected loss ratio.

Further, assume that the rating factors and indications underlying the prior and current reviews are as follows.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Data</td>
</tr>
<tr>
<td>(X)</td>
</tr>
<tr>
<td>Prior</td>
</tr>
<tr>
<td>Current</td>
</tr>
</tbody>
</table>

The increase in the indication from the prior to current review is \(\Delta I = 0.3538 - 0.0667 = 0.2871\), which is not unusual. The sequential replacement approach may proceed as follows: let \(\Delta I(X)\) be the impact due to the insurer's experience, \(X\). From equation (3), it follows that

\[
\Delta I(X) = \frac{1.0 \times 0.8 + (1.0 - 0.8) \times 0.4}{0.6} - \frac{0.7 \times 0.8 + (1.0 - 0.8) \times 0.4}{0.6} = 0.4.
\]

Let \(\Delta I(C)\) be the impact due to the credibility, \(C\). Then,

\[
\Delta I(C) = \frac{1.0 \times 0.7 + (1.0 - 0.7) \times 0.4}{0.6}
\]

\(\text{To keep this example simple, many of the rating factors, such as the on-level factor, trend, and loss development factor, are not considered in equation (3). The impacts of these factors will be discussed in detail in the next section.}\)
Let $\Delta I(B)$ be the impact due to the experience applied to the complement of credibility, $B$.

$$\Delta I(B) = \frac{1.0 \times 0.7 + (1.0 - 0.7) \times 0.6}{0.6}$$

$$\Delta I(B) = -0.1.$$ 

Finally, let $\Delta I(ELR)$ be the impact due to the expected loss ratio, $ELR$. Then,

$$\Delta I(ELR) = \frac{1.0 \times 0.7 + (1.0 - 0.7) \times 0.6}{0.65} - \frac{1.0 \times 0.7 + (1.0 - 0.7) \times 0.4}{0.6}$$

$$\Delta I(ELR) = -0.1128.$$ 

In the above calculations, the order of replacement is $X$ first, then $C$, then $B$, and finally $ELR$. If this order of replacement changes, however, the impact of each factor may change. For example, when the order of replacement is $ELR$ first, then $B$, then $C$, and finally $X$, we get

$$\Delta I(ELR) = -0.0821,$$ 

$$\Delta I(B) = 0.0615,$$ 

$$\Delta I(C) = -0.0154,$$ 

$$\Delta I(X) = 0.3231.$$ 

On the other hand, when the order of replacement is $B$ first, then $X$, then $ELR$ and finally $C$, we get

$$\Delta I(B) = 0.0667,$$ 

$$\Delta I(ELR) = -0.1179,$$ 

$$\Delta I(X) = 0.0400,$$ 

$$\Delta I(C) = -0.0615.$$ 

Given this problem with the sequential replacement approach, a new method is needed to compute the impact of each factor that is independent of the order of the computations. The chain rule approach described below solves this problem.
3 The Chain Rule Approach

3.1 Definition

Again, let $x_t = (x_1, x_2, \ldots, x_m)$ denote a vector of the $m$ factors used in determining the prior rate level indication at time $t$, and $f(x_t)$ be the rate level indication function. Consider what happens when there are infinitesimal changes in the rating factors. The total differential of the indication function can be calculated by the chain rule of differentiation (Edwards, 1973, Chapter 2):

$$df(x_t) = \sum_{i=1}^{m} \frac{\partial f(x_t)}{\partial x_{t,i}} dx_{t,i}. \tag{4}$$

Let $x_{t+1}$ be the current vector of rates. Then for small changes, however, equation (4) can be approximated by

$$\Delta f(x_t) = f(x_t + \Delta x_t) - f(x_t) \approx \sum_{i=1}^{m} \frac{\partial f(x_t)}{\partial x_{t,i}} \Delta x_{t,i} \tag{5}$$

where $\Delta x_t = x_{t+1} - x_t = (\Delta x_{t,1}, \Delta x_{t,2}, \ldots, \Delta x_{t,m})$. From equation (5), the individual impact of factor $i$ may be approximated by $[\frac{\partial f(x_t)}{\partial x_{t,i}}] \times dX_{t,i}$. Its marginal impact is approximated by $\frac{\partial f(x_t)}{\partial x_{t,i}}$. Note that this approach is not affected by the order of the estimation sequence.

In the real world, however, the chain rule approach has a serious limitation. For many real world applications, the changes in $x_t$ are not necessarily small so equation (5) cannot be used. To cope with a significant change in $x_t$, a multivariate Taylor series expansion can be used. Recall the multivariate Taylor series expansion:

$$\Delta f(x_t) = \sum_{i=1}^{m} \frac{\partial f(x_t)}{\partial x_{t,i}} \Delta x_{t,i} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 f(x_t)}{\partial x_{t,i} \partial x_{t,j}} \Delta x_{t,i} \Delta x_{t,j} + \cdots.$$ 

The chain rule can be approximated by the first order Taylor series expansion given in equation (5). This is only an approximation, however. But because we know the (exact) value of $\Delta f(x_t)$ we can make this approximation exact.

Now, by the mean value theorem, there is at least one point, $\chi$, given by

$$\chi = x_t + \tau \Delta x_t \text{ with } [0 \leq \tau \leq 1] \tag{6}$$

$f(x_t)$ is assumed to be at least twice differentiable in each of its parameters.
for which the first order Taylor series approximation is exact; see for example, Edwards (1973, Chapter 2). The theorem, however, does not indicate where \( x \) is. One obvious choice is to use the mid-point between \( x_t \) and \( x_{t+1} \) (i.e., at \( \tau = 0.5 \)) to evaluate the partial derivatives. As we will see, there is a better choice.

Consider the equation (as a function of \( \tau \)):

\[
H(\tau) = \Delta f(x_t) - \sum_{i=1}^{m} \left. \frac{\partial f(x_t)}{\partial x_{t,i}} \right|_{x_t=x_t^*} \times \Delta x_{t,i}
\]

(7)

where \( x \) is given in equation (6). Let \( \tau^* \) be the smallest value of \( \tau \) for which \( H(\tau) = 0 \), and let \( x_t^* \) be defined as

\[
x_t^* = x_t + \tau^* \Delta x_t.
\]

(8)

The mean value theorem only guarantees the existence of \( \tau^* \). We can determine \( \tau^* \) by first plotting \( H(\tau) \) for \( \tau = k/100 \), \( k = 1, 2, \ldots, 100 \) and observing the number and approximate location of the roots of \( H(\tau) \). Then \( \tau^* \) can be obtained more accurately using well-known numerical root-finding methods such as the bisection method or the secant method. (See, for example, Burden and Faires, 1985, Chapter 2.) In most practical situations, we expect \( \tau^* \) to be close to 0.5, i.e., \( \tau^* \approx 0.5 \).

The marginal impact and the impact of factor \( i \) can be defined as follows.

**Definition 1** Given a vector of \( m \) factors \( x_t = (x_1, x_2, \ldots, x_m) \) and the rate level indication function \( f(x_t) \), the marginal impact of factor \( i \), for \( i = 1, 2, \ldots, m \), is \( MIF(i) \) where

\[
MIF(i) = \left. \frac{\partial f(x_t)}{\partial x_{t,i}} \right|_{x_t=x_t^*}
\]

(9)

The impact of factor \( i \) can be defined as follows:

**Definition 2** The impact of factor \( i \), for \( i = 1, 2, \ldots, m \), is \( \Delta I(i) \) where

\[
\Delta I(i) = \left. \frac{\partial f(x_t)}{\partial x_{t,i}} \right|_{x_t=x_t^*} \times \Delta x_{t,i}
\]

\[
= MIF(i) \times \Delta x_{t,i}.
\]

(10)

3.2 Example 1 (Continued): The Chain Rule Approach

Let \( x_t = (X, C, B, ELR) \). Recall equation (3),

\[
I = f(x_t) = \frac{X \times C + (1 - C) \times B}{ELR} - 1.
\]
Clearly the partial derivatives are:

\[
\frac{\partial I}{\partial X} = \frac{C}{ELR}, \quad \frac{\partial I}{\partial C} = \frac{(X - B)}{ELR}, \quad \frac{\partial I}{\partial B} = \frac{(1 - C)}{ELR}, \quad \frac{\partial I}{\partial ELR} = -\left(\frac{X \times C + (1 - C) \times B}{ELR^2}\right).
\]

Given the data in Table 1, we have \(x_t = (0.7, 0.8, 0.4, 0.6)\) and \(\Delta x_t = (0.3, -0.1, 0.2, 0.05)\). Using equation (7), we have \(\tau^* = 0.4900\). Notice that, as expected, \(\tau^*\) is close to 0.5. From equation (8), \(x_t^* = x_t + 0.4900 \Delta x_t = (0.8470, 0.7510, 0.4980, 0.6245)\). Equations (9) and (10) now can be used to obtain the marginal impact and the impact of each factor. For example, the marginal impact of the factor \(ELR\) is

\[
MIF(ELR) = -\frac{0.8470 \times 0.7510 + (1 - 0.7510) \times 0.4980}{(0.6245)^2} = -1.9490.
\]

The impact of factor \(ELR\) is \(-1.9490 \times 0.05 = -0.0975\). Table 2 shows the marginal impact and the impact of each factor in this example.

<table>
<thead>
<tr>
<th>Impact of Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X)</td>
</tr>
<tr>
<td>(C)</td>
</tr>
<tr>
<td>(B)</td>
</tr>
<tr>
<td>(ELR)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

4 Example 2: Workers' Compensation Rating

4.1 The Problem

Next, an example from workers' compensation is considered. The example shows how to adjust the indication formula in order to consider the impacts of rating factors for trend, loss development, and any intervening rate changes.
Assume that we are proceeding on a state rate review in which the rate is stipulated by the Rating Bureau in that state. The insurer is free to use flexible rating tools, however, such as rate deviation, dividends, or schedule rating, to compete in the state.

Suppose the following information is given:

- The prior review uses the experience of 1990 accident year ending 12/31/90 evaluated as of 3/31/1992 (15 month maturity);
- The current review uses the experience of 1991 accident year ending 12/31/91 evaluated as of 3/31/1993 (15 month maturity);
- The Bureau's loss ratio is applied to the complement of credibility, and the prior and current reviews also use the 1991 and 1992 accident year experience, respectively, evaluated as of the same maturity date for the insurer's loss ratio;
- The target average effective date for the prior review is 7/1/1993;
- The target average effective date for the current review is 7/1/1994;
- An exponential trend with a 6 percent annual trend amount is used in the prior review for both insurer's and Bureau's loss ratios;
- An exponential trend with a 10 percent annual trend amount is used in the current review for both insurer's and Bureau's loss ratios; and
- There is a rate change of 15 percent between the two review periods.

The following loss ratio formula is used in this example to calculate the rate level indication:

\[ I = \frac{T \times F}{ELR} (X \times D \times L \times C + (1 - C) \times B) - 1 \]  

where \( I \) is the rate level indication; \( X \) is the insurer's on-leveled but untrended and undeveloped loss ratio; \( D \) is the loss development factor; \( L \) is the loss adjustment expense factor; \( C \) is the credibility; \( B \) is the untrended Bureau loss ratio; \( T \) is the trend factor; \( F \) is the flexible rating factor (such as rate deviation and schedule rating); and \( ELR \) is the expected loss ratio.

Table 3 lists all the values assumed for these rating factors in the two reviews and the resulting prior and current review indications.
Both reviews use an exponential trend, but with different annual trend amounts: 6 percent for the prior review and 10 percent for the current review. The trending period for both reviews is the same, three years: from 7/1/90 to 7/1/93 for the prior review and from 7/1/91 to 7/1/94 for the current review. Thus, the trend factor in the prior review was $(1.06)^3 = 1.1910$, while in the current review it is $(1.10)^3 = 1.331$. In addition, the overall indication change is

$$
\Delta I = 0.3159 - 0.0946 = 0.2213.
$$

Before applying the chain rule approach, several adjustments must be made to the indication formula given in equation (11). Adjustments are made to the following factors: rate on-level, trend, and loss development. This is because these rating factors must be compared at the same point in time between the two reviews. Thus, adjustments are necessary if there have been any rate changes between reviews, different trends are selected, or if the experience is evaluated on different maturity dates.

One rating factor not considered in this example is the benefit changes between reviews. Similar to the rate change, the insurer's loss ratio and the Bureau's loss ratio reflect all benefit changes through the reviews. Therefore, the adjustment for benefit change will impact the formula in essentially the same way as the adjustment for rate change, as discussed below.
4.2 Adjustment for Rate Change

The insurer's loss ratios and the Bureau's loss ratios listed in Table 3 reflect all rate changes through each review. Because there was a 15 percent rate change between the two review periods, the loss ratios are inconsistent. One way to adjust for the rate change is to recalculate the loss ratios in the current review without the 15 percent rate change and add one more rating factor, $R$, for the rate change to equation (11). Let $X'$ be the insurer's adjusted loss ratio and $B'$ be the Bureau's adjusted loss ratio. Table 4 shows their values.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Adjustment for Rate Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors</td>
<td>Prior</td>
</tr>
<tr>
<td>$X'$</td>
<td>0.4200</td>
</tr>
<tr>
<td>$B'$</td>
<td>0.6300</td>
</tr>
<tr>
<td>$R$</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Note: $0.4400 \times 1.1500 = 0.5060$ and $0.6500 \times 1.1500 = 0.7475$

4.3 Adjustment for Trend

The impact of trend on the rate indication can be split into two parts: the impact due to the trend amount and the impact due to the trend date. In this example, the annual trend amounts are different between the two reviews: 6 percent in the prior review and 10 percent in the current review. Also, the impact of the trend date must be evaluated separately because a more recent review will trend the on-level experience into a later effective date, which is one year later in this example. The trend date impact represents the increase in costs from the prior target average effective date to the current target average effective date.

The overall trend impact can be broken into the trend amount impact and the trend duration impact as follows: The average accident date (7/1/90) of the experience period in the prior review is used as the point in time to compare the trend impact between the two reviews. First the insurer's loss ratio and Bureau's loss ratio in the current review are detrended backward from 7/1/91 to 7/1/90 using the 10 percent trend amount. Next the trend amount impact for both reviews is defined from 7/1/90 to 7/1/93, which is $1.10^3 = 1.331$ for the current review and $1.06^3 = 1.191$ for the prior review. The difference between these two numbers is due to the different trend amounts used. Because the experience in the current review is trended one year beyond
the prior review (from 7/1/93 to 7/1/94) the trend date impact for the current review is defined as 1.10, while the trend date impact for the prior review is assumed to be 1.0. The trend date impact reflects the loss cost inflation from the prior target date to the current target date.

Following the previous adjustment for the rate change in Table 4, we further adjust the indication formula for the trend impact as follows: let TA be the trend amount factor and TD be the trend date factor, then

<table>
<thead>
<tr>
<th>Factors</th>
<th>Prior</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X''$</td>
<td>0.420</td>
<td>0.4600</td>
</tr>
<tr>
<td>$B''$</td>
<td>0.630</td>
<td>0.6795</td>
</tr>
<tr>
<td>$TA$</td>
<td>1.191</td>
<td>1.3310</td>
</tr>
<tr>
<td>$TD$</td>
<td>1.000</td>
<td>1.1000</td>
</tr>
</tbody>
</table>

Note: $0.5060/1.10 = 0.4600$, and $0.7475/1.10 = 0.6795$.

4.4 Adjustment for Loss Development Factor

In addition to the adjustments for rate change and trend, we need to ensure that the loss data in the prior and current reviews are evaluated as of the same maturity date. That is, $X''$ and $D$ must represent the experience and development factor of the same maturity between the two reviews. If not, an adjustment must be made to one of the reviews so that the two reviews are consistent.

For example, assume that prior review data are 12 months matured, while current review data are 15 months matured. We can make an adjustment to the prior review by dividing the prior 12-to-ultimate factor into a 12-to-15 factor and a 15-ultimate factor. Then the prior experience is combined with the 12-to-15 factor. By doing so, the loss experience and development factors between the two reviews become comparable. In this workers' compensation example, however, the insurer's loss ratio and the Bureau's loss ratio between the two reviews are developed from the same maturity date to ultimate; thus, there is no need for this adjustment.

4.5 Application of the Chain Rule Approach

At this point, we have finished all the necessary adjustments, and we are ready to adjust equation (11) to reflect all of the adjustments.
made thus far.

\[
I = \left[ \frac{TA \times TD \times F}{R \times ELR} \right] \left( X'' \times D \times L \times C + (1 - C) \times B'' \right) - 1. \tag{12}
\]

Table 6 summarizes the prior data and the current (adjusted) data needed for equation (12). It directly gives us \( x_t \) and \( \Delta x_t \). From the equations for the partial derivatives, we can calculate \( \tau^* \) and hence \( x_t^* \):

\[
x_t = (0.42, 1.35, 1.15, 0.85, 0.63, 1.191, 1.0, 1.02, 1.0, 0.72)
\]

\[
\Delta x_t = (0.04, 0.025, -0.002, 0.05, 0.0495, 0.14, 0.1, -0.01, 0.15, -0.015)
\]

\[
\tau^* = 0.50376
\]

\[
x_t^* = (0.4402, 1.3626, 1.1490, 0.8752, 0.6550, 1.2615, 1.0504, 1.0150, 1.0756, 0.7124)
\]

### Table 6

**Adjusted Review Data**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Prior</th>
<th>Current</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X'' )</td>
<td>0.4200</td>
<td>0.4600</td>
<td>0.0400</td>
</tr>
<tr>
<td>( D )</td>
<td>1.3500</td>
<td>1.3750</td>
<td>0.0250</td>
</tr>
<tr>
<td>( L )</td>
<td>1.1500</td>
<td>1.1480</td>
<td>-0.0020</td>
</tr>
<tr>
<td>( C )</td>
<td>0.8500</td>
<td>0.9000</td>
<td>0.0500</td>
</tr>
<tr>
<td>( B'' )</td>
<td>0.6300</td>
<td>0.6795</td>
<td>0.0495</td>
</tr>
<tr>
<td>( TA )</td>
<td>1.1910</td>
<td>1.3310</td>
<td>0.1400</td>
</tr>
<tr>
<td>( TD )</td>
<td>1.0000</td>
<td>1.1000</td>
<td>0.1000</td>
</tr>
<tr>
<td>( F )</td>
<td>1.0200</td>
<td>1.0100</td>
<td>-0.0100</td>
</tr>
<tr>
<td>( R )</td>
<td>1.0000</td>
<td>1.1500</td>
<td>0.1500</td>
</tr>
<tr>
<td>( ELR )</td>
<td>0.7200</td>
<td>0.7050</td>
<td>-0.0150</td>
</tr>
<tr>
<td>( I )</td>
<td>0.0946</td>
<td>0.3159</td>
<td>0.2213</td>
</tr>
</tbody>
</table>
Table 7

<table>
<thead>
<tr>
<th>Factors</th>
<th>MIF</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X''$</td>
<td>2.4049</td>
<td>0.0962</td>
</tr>
<tr>
<td>$D$</td>
<td>0.7768</td>
<td>0.0194</td>
</tr>
<tr>
<td>$L$</td>
<td>0.9212</td>
<td>-0.0018</td>
</tr>
<tr>
<td>$C$</td>
<td>0.0599</td>
<td>0.0030</td>
</tr>
<tr>
<td>$B''$</td>
<td>0.2191</td>
<td>0.0108</td>
</tr>
<tr>
<td>$TA$</td>
<td>0.9528</td>
<td>0.1334</td>
</tr>
<tr>
<td>$TD$</td>
<td>1.1443</td>
<td>0.1144</td>
</tr>
<tr>
<td>$F$</td>
<td>1.1843</td>
<td>-0.0118</td>
</tr>
<tr>
<td>$R$</td>
<td>-1.1175</td>
<td>-0.1676</td>
</tr>
<tr>
<td>$ELR$</td>
<td>-1.6871</td>
<td>0.0253</td>
</tr>
<tr>
<td>Total</td>
<td>0.2213</td>
<td></td>
</tr>
</tbody>
</table>

5 Summary

The chain rule approach has been introduced in this paper to reconcile two rate level indications that have been made at different rate review periods. This approach individually estimates the impact of each rating factor on the overall indication change. Unlike the sequential replacement approach, the chain rule approach does not depend on a sequence of estimation. This paper further indicates evaluating partial derivatives at the mid-point between the prior and current reviews provide a close approximation to the overall indication change. A workers' compensation example is given to show how to adjust the rate level indication formula for trend, loss development, and any rate and benefit changes between two reviews.

Although the main body of the discussion focuses on the loss ratio method, the developed chain rule approach can be applied equally to the pure premium method, such as the pure premium formula noted by McClenahan (1990, Chapter 2):

$$RT = (PP + FE)/(1 - VE)$$

where $RT$ is the indicated rate per unit of exposure; $PP$ is the trended and developed pure premium per unit of exposure; $FE$ is the fixed expense per unit of exposure; and $VE$ is the variable expense per unit of exposure.
While the loss ratio method develops the indicated percent change in the rate, the pure premium method develops the indicated rate. The $PP$ term in the above formula can be subdivided into loss development and trend factors. The subsequent procedure to estimate the impact of each factor on the change in the indicated rate remains the same as described earlier in this paper.

References


Tax Assistance to Qualified Retirement Savings Plans: Deferral or Waiver?

Robert L. Brown*

Abstract

There exist significant tax incentives for retirement savings plans in Canada and the United States. Qualified employer and employee contributions, within limits, are tax deductible to the employer and nontaxable to the employee. Also, investment income is not taxed until taken. On the other hand, monies received from funds having such tax incentives are taxable in full as income to the recipient when taken. This paper analyzes the two tax advantages of qualified retirement savings plans: the tax deductibility of contributions and the nontaxation of investment income until it has been distributed. The algebraic analysis shows that the deductibility of contributions represents a deferral of tax, but that it does not create any permanent loss of revenue to the government. On the other hand, the algebra indicates that there is a permanent tax subsidy associated with the deferred taxation of investment income.

Key words and phrases: tax deductions, savings vehicle, contributions, accumulated value

1 Introduction

Canadian and United States laws provide significant tax incentives for individuals to save for retirement through qualified vehicles. There are two tax incentives provided in the United States and Canada.

*Robert L. Brown, F.C.I.A., F.S.A., A.C.A.S., is professor of statistics and actuarial science and director of the Institute of Insurance and Pension Research at the University of Waterloo. He is a past president of the Canadian Institute of Actuaries and is currently on the Society of Actuaries' Board of Governors and Executive Committee. He is also an elected Councillor in the City of Waterloo. Professor Brown has authored several articles and books.

Professor Brown's address is: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON N2L 3G1, Canada.
1. Employer contributions to qualified plans are tax deductible to the employer and nontaxable to the employee. For employees and individuals saving for retirement through qualified vehicles, their contributions, within limits, are also tax deductible (e.g., IRAs, 401(k) plans in the United States and RRSPs in Canada\(^1\)).

2. For these qualified plans, the investment income earned on the pension funds is not taxable until it is paid out. Income derived from these funds, however, is fully taxable to the individual who receives it.

What is the value of these tax incentives? In particular, do these incentives effectively result in deferred taxes, or is the outcome a waiver of taxes?\(^2\)

It often is stated that these incentives represent only tax deferral and are not a tax expenditure or permanent tax subsidy. For example, Johansen (1993) states:

> But when the plan starts paying out benefits, the recipients will have to pay the appropriate income tax on those benefits. So the tax-exempt status of qualified pension plans creates a tax deferral—not a tax expenditure.

Similarly, in a discussion of Aitken's (1991) paper that claims there is a permanent tax subsidy implicit in the nontaxation of the annual investment income earnings, Flanagan (1991) states:

> One does not need to be an actuary to realize that the author's fundamental point is flawed. There is tax on the investment income accumulating in a registered plan, but the tax on the investment income, like the tax on the principal, is deferred until the payout period.

The objective of this paper is to review the two tax incentives (cited earlier) that are provided to retirement savings vehicles and to determine algebraically whether such incentives are essentially tax deferrals or if they result in a tax waiver. The paper also will present a summary of the tax advantages associated with alternative savings vehicles. It is

\(^1\)This is not meant to be an exhaustive list. Any plan with these tax advantages is meant to be included, such as some profit sharing plans.

\(^2\)In this paper, the term \textit{tax deferral} means that for that particular tax provision the accumulated value of the taxes paid is the same with or without the provision. Note that the deferral still may be viewed as advantageous. If the accumulated value of the taxes paid with the provision is smaller than that paid without the provision, however, then the provision results in a \textit{tax waiver}.
well known (and obvious) that for persons who expect to be in a lower tax bracket after retirement than before retirement (which often is expected), there are permanent tax advantages to using qualified savings vehicles to save for retirement. Thus, this paper will not investigate that particular aspect of the tax advantages.

2 Advantage of Alternative Savings Vehicles

What are the tax advantages associated with the ability to take a tax deduction for contributions made to a qualified vehicle? To explore this issue, the following notation is needed: $T$ is the marginal tax rate; $I$ is the gross investment rate of return (for all investments) per annum; $i$ is the net rate of return per annum; $C$ is the before tax contribution; and $n$ is the time from contribution to withdrawal.

To simplify the presentation, the following assumptions are made:

1. $T$, $I$, $C$, and $i$ are constant throughout the period under consideration, $n$ years. In addition, they do not vary by whether the fund is a qualified fund or not or whether the fund is private or public.

2. The marginal tax rate is the same before and after retirement.

3. The value of a tax incentive is defined to be the difference between the accumulated value of certain defined contributions without and with the tax incentive.³

2.1 Tax Deductibility of Contributions

What advantage is gained purely from the tax deductibility of contributions? To determine this advantage, it will be assumed that the rate of return on the funds is the after-tax rate, so

$$i = I \times (1 - T).$$

Table 1 shows that the after-tax accumulated incomes for qualified and nonqualified vehicles are equal (ignoring the effects of taxes on investment income).

³Further possible investment or expenditure considerations are beyond the scope of the illustrations contained herein.
It is clear that the tax advantage associated with the deductibility of contributions is purely an advantage of tax deferral. If one's marginal tax rate is the same before and after retirement, then there is no permanent tax waiver or tax subsidy associated with the deductibility of contributions.

2.2 Nontaxation of Investment Income

Within the qualified vehicle, funds grow at a rate of $I$ per annum. Income derived from these funds is taxed at the marginal rate, $T$, when disbursed. Within the nonqualified plan, funds will grow at rate $i = I(1 - T)$ per annum, but funds will not be taxed when taken out.

Again, consider a before-tax contribution of $C$ within either a qualified or nonqualified vehicle. For the qualified plan, the net receipt to the retiree is $C(1 - T)(1 + I)^n$, while for the nonqualified plan it is $C(1 - T)(1 + i)^n$. One must remember that the tax deductibility of the contribution provides no net gain and explains none of the difference between the two values above. Thus, the gain represented by the difference of the two values above can be categorized as coming from the difference in the taxation of investment income. That gain is:

$$C(1 + I)^n(1 - T) - C(1 - T)(1 + i)^n = C(1 - T)[(1 + I)^n - (1 + i)^n].$$

As $i = I(1 - T)$, it follows that $i < I$; there must be a net gain. A numerical example illustrates these points. You are given the following information:

- Before-tax contribution: $C = $2,000
- Marginal tax rate: $T = 40$
- Gross rate of return per annum: $I = 7$
- Net rate of return per annum: $i = 0.07(1 - 0.40) = 4.2$
- Time from contribution to withdrawal: $n = 30$ years.
Using a qualified vehicle, the retiree receives:

\[ \$2,000(1.07)^{30}(1 - 0.40) = \$9,134.71. \]

On the other hand, using a nonqualified vehicle yields the retiree:

\[ \$2,000(1 - 0.40)(1.042)^{30} = \$4,123.00. \]

The net gain to the retiree by using the qualified fund is \$5,011.71. But what is the source of this \$5,011.71 gain?

One must conclude that the \$5,011.71 comes from a direct tax waiver or subsidy. To prove this assertion, look at the tax revenues that accrue in each situation. For the qualified fund, the government gets

\[ C \times T \times (1 + I)^n = \$6,089.80 \text{ at time } t = 30. \]

In the nonqualified fund, however, the government gets \( C \times T = \$800 \) immediately which, at time \( t = 30 \), is worth:

\[ C \times T \times (1 + I)^n = \$800(1.07)^{30} = \$6,089.80. \]

Thus, as proven before, there is no tax waiver or subsidy associated with the tax deductibility of contributions, only tax deferral. Under the nonqualified fund, however, the government receives additional taxes: the taxes on the yearly investment income on the fund. In this example, the accumulated value of this tax on annual investment income at time \( t = 30 \) is:

\[ C \times (1 - T) \times T \times I \times \sum_{k=0}^{t-1} (1 + i)^k(1 + I)^{t-1-k} = \$5011.71. \]

That is, the gain to the retiree who uses a qualified fund is equal to the permanent tax revenue loss to the government under the assumptions given.

The nontaxation of the investment income on the qualified fund until taken as income clearly is a permanent tax waiver, not a tax deferral.

3 Extensions

The expressions for the tax impact on qualified pension funds, derived in Section 2, can be adjusted to include other insurance and savings vehicles. The table below presents the tax effects in summary form.
Table 2

The Effects of Taxes on Various Vehicles

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Taxes?</th>
<th>Frequency of Taxes</th>
<th>Rate of Taxation</th>
<th>After-tax Accumulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF</td>
<td>Yes</td>
<td>Annually</td>
<td>C-Gains</td>
<td>([1 + I(1 - gT)]^n)</td>
</tr>
<tr>
<td>MMF</td>
<td>Yes</td>
<td>Annually</td>
<td>Ordinary</td>
<td>([1 + I(1 - T)]^n)</td>
</tr>
<tr>
<td>IP</td>
<td>Yes</td>
<td>Never</td>
<td>Exempt</td>
<td>((1 + I)^n)</td>
</tr>
<tr>
<td>SPDA</td>
<td>Yes</td>
<td>Deferred</td>
<td>Ordinary</td>
<td>((1 + I)^n(1 - T) + T)</td>
</tr>
<tr>
<td>PF</td>
<td>No</td>
<td>Deferred</td>
<td>Ordinary</td>
<td>((1 + I)^n)</td>
</tr>
<tr>
<td>FC</td>
<td>Yes</td>
<td>Deferred</td>
<td>C-Gains</td>
<td>((1 + I)^n(1 - gT) + gT)</td>
</tr>
</tbody>
</table>

MF = Mutual Funds; MMF = Money Market Funds; IP = Insurance Policies; SPDA = Single Premium Deferred Annuities; PF = Pension Funds; FC = Foreign Corporations; and C-Gains = Capital Gains; and \(gT\) = The capital gains tax rate.

The following is a brief description of the various savings vehicles.

- **Money Market Funds (MMF)**: This vehicle is the ordinary savings account. Deposits are not tax deductible, and investment income normally is taxed fully each year at ordinary tax rates. This is the least advantageous of the savings vehicles.

- **Mutual Funds (MF)**: These refer to those mutual funds that are not money market funds. Here deposits are not tax deductible. Investment income is taxed at the capital gains tax rate, however, which is given in the table as rate \(gT\). In the United States current tax rates for capital gains are subject to a 28 percent limitation, while there is no such limitation on ordinary income. Dividend and interest received by the mutual fund and capital gains realized by the mutual fund are taxable to shareholders annually.

- **Insurance Policies (IP)**: This category refers to those insurance policies that have achieved exempt status. While deposits are not tax deductible, the earnings on the investment are entirely tax exempt. The gain on disposition is taxable in the hands of policyholders unless the proceeds are paid as a death benefit. Further, the insurance company pays some tax on its investment income.

- **Single Premium Deferred Annuities (SPDA)**: Deposits are not tax deductible, but the taxes on the investment income are deferred
until the policyholder takes the money out as income. The same applies to IRA contributions that are not deductible because the owners have income above certain limits specified by law. The value of this deferral is the same as in the qualified pension plan. The Tax Reform Act of 1986 in the United States eliminated the ability of corporations and partnerships to defer tax with single premium deferred annuities. Only individual investors can use SPDAs to defer tax on the investment income. Also there exists an excise tax of 10 percent if the SPDA is surrendered, in whole or in part, prior to age 59.5 unless the withdrawals take the form of a life annuity. Finally, the insurance company pays some tax on its investment income.

- **Pension Funds (PF):** The tax advantages of qualified pension funds have been discussed in detail previously. When tax rates are constant over time, insurance policies that are tax exempt are equivalent to pension funds that are qualified.

- **Foreign Corporations (FC):** Again, deposits are not tax deductible; however, the tax on the earnings is deferred and taxed at capital gains rates when the investment is liquidated. Examples include an investment in the common stock of an investment company located in a tax haven or bond investments held by corporations in tax havens.4

When \( g = 0 \), mutual funds, foreign corporations, insurance policies, and pension funds are equivalent vehicles. When \( g = 1 \), investments in foreign corporations and single premium deferred annuities are equivalent.

In general, it is more accurate to list the accumulated value of the dollar invested in the qualified pension fund as \[ (1 - T_n)/(1 - T_0) \] where \( T_n \) and \( T_0 \) represent the marginal tax rates at the time of contribution \( t = 0 \) and at the time of withdrawal \( t = n \). This paper assumes that these two tax rates are the same. But one would expect the marginal rate \( T_n \) to be slightly less than \( T_0 \) which, as mentioned previously, provides a further tax advantage.

4A tax haven is a country or other political entity that offers outside businesses and individuals a climate of minimal or nonexistent taxation. In some cases, the low taxes apply not only to those levied by the tax haven itself, but also to the possibility of reducing or avoiding taxes levied in the investor's home country (Scott, 1988, p. 353).
4 Conclusions

This paper has looked at the tax incentives provided in several savings vehicles and qualified pension funds in particular. The paper has shown that the tax advantage associated with the deductibility of tax contributions is one of tax deferral, but not tax avoidance or permanent tax waiver. On the other hand, the paper shows that the tax advantage associated with the nontaxation of investment income on qualified funds until taken is a tax waiver or tax subsidy from the government to participants of qualified plans.

Further public policy debate on the impact of tax concessions is needed. The author hopes that this paper will spark such a debate and assist in an intelligent discussion.

References


