1993

Journal of Actuarial Practice, Volume 1, No. 2, 1993

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

Follow this and additional works at: 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.
ARTICLES

Ethical Issues, Helps, and Challenges:
Perceptions of U.S. Actuaries
Therese M. Vaughan, Robert W. Cooper, and Garry L. Frank................. 5

Managing the Relative Volumes of Participating and Nonparticipating
Business in a Mutual Life Company
Robert G. Chadburn.................................................................................... 27

A Critique of Defined Contribution
Plans Using a Simulation Approach
David M. Knox............................................................................................ 49

Discussion
Michael Sze..................................................................................................... 67

David M. Knox's Reply............................................................................ 68

The Definition of Insurance:
Implications for a Health Insurance Demand Model
Mark J. Browne ............................................................................................ 71

Discussion
Charles S. Fuhrer.......................................................................................... 93

Mark J. Browne's Reply.............................................................................. 94

On the Equivalence of the Loss Ratio and Pure Premium Methods of
Determining Property and Casualty Rating Relativities
Robert L. Brown.......................................................................................... 97

Funding Methods and Pension Plan Amendments
Keith P. Sharp.............................................................................................. 111

Estimating the Effect of Statutory Changes on Insured Casualty Losses
Using Generalized Indicator Variables
Ruy A. Cardoso............................................................................................ 127

Life Insurance Applications of Recursive Formulas
L. Timothy Giles............................................................................................ 141
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. One of the goals of the Journal of Actuarial Practice is to improve communication between the practicing and the academic actuarial communities. In addition, the Journal provides a forum for the presentation and discussion of ideas, issues (controversial or otherwise), and methods of interest to actuaries.

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:

- 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 unconditionally, (2) accept subject to minor revisions, (3) accept contingent on substantive revisions, (4) resubmit, return to the author(s) for major revisions and subsequent resubmission, or (5) reject.

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

See back cover for instructions to authors.
Ethical Issues, Helps, and Challenges: Perceptions of U.S. Actuaries

Therese M. Vaughan, Robert W. Cooper, and Garry L. Frank*

Abstract

This paper reports the findings of a survey of Fellows of the Casualty Actuarial Society (FCAS) to determine their perceptions of the key ethical issues and dilemmas facing the industry today and the factors they view as most helpful and challenging in resolving these dilemmas. The responses are compared to a previous survey of Fellows of the Society of Actuaries (FSA). The study finds that FSAs and FCASs tend to rank key ethical issues similarly and that both groups of actuaries tend to look first to their own personal values, second to certain factors in their business environment, and last to professional factors when resolving ethical dilemmas. Finally, the paper contains some implications for the actuarial professional associations as they attempt to assist their members in resolving ethical dilemmas.

Key words: ethics, professionalism

1 Introduction

Recent years have seen significant activity among American actuarial organizations focused on professionalism. Both the Society of Actuaries (SOA) and the Casualty Actuarial Society (CAS) have instituted admissions courses that include professionalism and ethics.1 More recently the organizations have cooperated through the American Academy of Actuaries (AAA) to create the Actuarial Board for Counseling and Discipline (ABCD), from which all actuar-

* Therese M. Vaughan, Ph.D., ASA, ACAS, CPCU, is director of the Insurance Center at Drake University where she teaches courses in insurance and actuarial science.

Robert W. Cooper, Ph.D., is Employers Mutual Distinguished Professor of Insurance at Drake University. He previously was dean at The American College where he was responsible for the CLU and ChFC designation programs.

Garry L. Frank, Ph.D. is a professor of public administration at Drake University where he teaches a course in business ethics.

1 In the case of the CAS, this course must be completed prior to associateship; the SOA course is required for admission to fellowship.
ies are encouraged to seek help in matters dealing with professional conduct. All three organizations have promulgated codes of professional conduct to encourage professionalism and ethical behavior by their members.

This paper reports the findings of two surveys of Fellows of the Casualty Actuarial Society eliciting their views on the major ethical problems today in the property/liability insurance industry and the factors they find helpful and challenging personally in resolving the ethical dilemmas they face at work. The surveys replicate previous surveys of Fellows of the Society of Actuaries conducted by Cooper and Frank (1992a, 1992b) in February 1991. A comparison of the results across actuarial groups suggests that the most problematic ethical issues are similar, as are the factors viewed as most helpful and challenging. By focusing on the issues of greatest concern and the factors viewed as most helpful in resolving them, some insights can be gained into approaches that may be taken by actuarial organizations to assist their members.

2 Methodology

2.1 The Survey Forms

Two survey instruments are used. One measures the perceptions of Fellows of the Casualty Actuarial Society with respect to ethical issues and dilemmas important in the property/liability insurance industry. The other measures the factors that are helpful and challenging in resolving ethical dilemmas. The first survey contains an itemized list of 34 potential ethical issues (dilemmas); they are presented in Table 1. Issues 8 and 9 are most directly related to the problems of selecting assumptions to use in pricing and reserving and those encountered when dealing with regulators. The remaining 28 of the first 30 issues reflect ethical issues and dilemmas facing businesses and their employees in general. Issues 31 through 34 deal with ethical dilemmas of particular concern to business professionals. Survey participants are asked to rate each of the 34 statements on a five point scale, where a 5.0 means that it is a major ethical problem in the property/liability insurance industry and a 1.0 means that it is not a problem. In addition to rating the 34 issues presented in Table 1, respondents are asked to indicate what they feel is the most important specific ethics problem facing those who work in the industry today. Finally, survey participants are asked to provide some demographic information, including how many years they have been mem-
<table>
<thead>
<tr>
<th>Issue</th>
<th>Description</th>
<th>FCAS—Employees</th>
<th>FCAS—Consultants</th>
<th>FSA—Employees</th>
<th>FSA—Consultants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
<td>Rank</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.667</td>
<td>2</td>
<td>2.947</td>
<td>3</td>
</tr>
<tr>
<td>Issue 1</td>
<td>Failure to provide products and services of the highest quality in the eyes of the customer</td>
<td>2.488</td>
<td>6</td>
<td>2.827</td>
<td>5</td>
</tr>
<tr>
<td>Issue 2</td>
<td>Failure to provide prompt, honest responses to customer inquiries and requests</td>
<td>1.940</td>
<td>16</td>
<td>1.967</td>
<td>21</td>
</tr>
<tr>
<td>Issue 3</td>
<td>Making disparaging remarks about competitors, their products, or their employees or agents</td>
<td>1.810</td>
<td>23</td>
<td>2.079</td>
<td>19</td>
</tr>
<tr>
<td>Issue 4</td>
<td>Misuse of proprietary information</td>
<td>1.786</td>
<td>24</td>
<td>2.053</td>
<td>20</td>
</tr>
<tr>
<td>Issue 5</td>
<td>Misuse of sensitive information belonging to others</td>
<td>1.643</td>
<td>27</td>
<td>1.882</td>
<td>25</td>
</tr>
<tr>
<td>Issue 6</td>
<td>Improper methods of gathering competitors' information</td>
<td>2.309</td>
<td>10</td>
<td>2.553</td>
<td>9</td>
</tr>
<tr>
<td>Issue 7</td>
<td>False or misleading representation of products or services in marketing, advertising, or sales efforts</td>
<td>3.277</td>
<td>1</td>
<td>3.592</td>
<td>1</td>
</tr>
<tr>
<td>Issue 8</td>
<td>Responding to pressure from clients and/or management to change assumptions used in pricing or reserving</td>
<td>2.512</td>
<td>4</td>
<td>2.829</td>
<td>4</td>
</tr>
<tr>
<td>Issue 9</td>
<td>Misrepresenting information provided to regulators</td>
<td>2.512</td>
<td>4</td>
<td>2.829</td>
<td>4</td>
</tr>
<tr>
<td>Issue</td>
<td>Problem Description</td>
<td>FCAS—Employees</td>
<td>FCAS—Consultants</td>
<td>FSA—Employees</td>
<td>FSA—Consultants</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>--------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
</tr>
<tr>
<td>Issue 10</td>
<td>Conflicts between opportunities for personal financial gain (or other personal benefits) and proper performance of one's responsibilities</td>
<td>2.250</td>
<td>11</td>
<td>2.671*</td>
<td>8</td>
</tr>
<tr>
<td>Issue 11</td>
<td>Conflicts of interest involving business or financial relationships with customers, suppliers, or competitors that influence or appear to influence one's ability to perform his or her responsibilities</td>
<td>1.929</td>
<td>17</td>
<td>2.453*</td>
<td>11</td>
</tr>
<tr>
<td>Issue 12</td>
<td>Conflicts of interest involving the marketing of products or services competing with those of one's own company</td>
<td>1.560</td>
<td>30</td>
<td>1.865*</td>
<td>27</td>
</tr>
<tr>
<td>Issue 13</td>
<td>Conflicts of interest that involve working for a competitor, customer, or supplier without approval</td>
<td>1.393</td>
<td>33</td>
<td>1.750*</td>
<td>32</td>
</tr>
<tr>
<td>Issue 14</td>
<td>Misuse of company assets/property</td>
<td>1.869</td>
<td>21</td>
<td>1.974*</td>
<td>22</td>
</tr>
<tr>
<td>Issue 15</td>
<td>Insider trading/other security trading problems</td>
<td>1.714</td>
<td>25</td>
<td>1.836*</td>
<td>28</td>
</tr>
<tr>
<td>Issue 16</td>
<td>Giving excessive gifts or entertainment</td>
<td>1.643</td>
<td>27</td>
<td>1.776*</td>
<td>31</td>
</tr>
<tr>
<td>Issue 17</td>
<td>Receiving excessive gifts or entertainment</td>
<td>1.607</td>
<td>29</td>
<td>1.671*</td>
<td>33</td>
</tr>
<tr>
<td>Issue</td>
<td>FCAS—Employees</td>
<td>FCAS—Consultants</td>
<td>FSA—Employees</td>
<td>FSA—Consultants</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
</tr>
<tr>
<td>Issue 18</td>
<td>Offering or soliciting payments or contributions for the purpose of influencing customers or suppliers</td>
<td>1.464</td>
<td>32</td>
<td>1.790•</td>
<td>29</td>
</tr>
<tr>
<td>Issue 19</td>
<td>Offering or soliciting payments or contributions for the purpose of influencing government officials</td>
<td>1.560</td>
<td>30</td>
<td>1.787</td>
<td>30</td>
</tr>
<tr>
<td>Issue 20</td>
<td>Offering or soliciting payments or contributions for the purpose of obtaining, giving, or keeping business</td>
<td>1.679</td>
<td>26</td>
<td>1.947•</td>
<td>24</td>
</tr>
<tr>
<td>Issue 21</td>
<td>Offering or soliciting payments or contributions for the purpose of persuading employees of another company to fail to perform or improperly perform their duties</td>
<td>1.214</td>
<td>34</td>
<td>1.250</td>
<td>34</td>
</tr>
<tr>
<td>Issue 22</td>
<td>Offering or soliciting payments or contributions for the purpose of influencing legislation or regulations</td>
<td>1.821</td>
<td>22</td>
<td>2.171•</td>
<td>15</td>
</tr>
<tr>
<td>Issue 23</td>
<td>Willful inaccuracy of books, records, or reports</td>
<td>2.000</td>
<td>15</td>
<td>2.342•</td>
<td>12</td>
</tr>
<tr>
<td>Issue 24</td>
<td>Abuse of expense accounts</td>
<td>2.048</td>
<td>14</td>
<td>2.227</td>
<td>14</td>
</tr>
<tr>
<td>Issue 25</td>
<td>Anti-trust issues</td>
<td>2.202</td>
<td>12</td>
<td>2.171</td>
<td>15</td>
</tr>
<tr>
<td>Issue 26</td>
<td>Relations with local communities</td>
<td>1.905</td>
<td>19</td>
<td>2.110</td>
<td>18</td>
</tr>
<tr>
<td>Issue</td>
<td>Description</td>
<td>FCAS—Employees</td>
<td>FCAS—Consultants</td>
<td>FSA—Employees</td>
<td>FSA—Consultants</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>---------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Rating</td>
<td>Rank</td>
<td>Mean Rating</td>
<td>Rank</td>
</tr>
<tr>
<td>27</td>
<td>Office/agency closings and layoffs</td>
<td>2.452</td>
<td>7</td>
<td>2.160</td>
<td>17</td>
</tr>
<tr>
<td>28</td>
<td>Discrimination in the workplace</td>
<td>2.179</td>
<td>13</td>
<td>2.289</td>
<td>13</td>
</tr>
<tr>
<td>29</td>
<td>Drug and alcohol abuse</td>
<td>1.893</td>
<td>20</td>
<td>1.973</td>
<td>23</td>
</tr>
<tr>
<td>30</td>
<td>Employee theft</td>
<td>1.917</td>
<td>18</td>
<td>1.882</td>
<td>25</td>
</tr>
<tr>
<td>31</td>
<td>Lack of knowledge or skills to perform one's duties competently</td>
<td>2.512</td>
<td>4</td>
<td>2.789</td>
<td>6</td>
</tr>
<tr>
<td>32</td>
<td>Failure to identify the customer's needs and recommend products and services</td>
<td>2.578</td>
<td>3</td>
<td>2.960*</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>Failure to be objective with others in one's business dealings</td>
<td>2.325</td>
<td>9</td>
<td>2.553</td>
<td>9</td>
</tr>
<tr>
<td>34</td>
<td>Misrepresenting or concealing limitations in one's abilities to provide</td>
<td>2.441</td>
<td>8</td>
<td>2.790*</td>
<td>6</td>
</tr>
</tbody>
</table>

* = significantly greater than the FCAS—employee value at the 0.05 level
** = significantly greater than the FSA—consultant value at the 0.05 level
# = significantly greater than the FCAS—consultant value at the 0.05 level
## = significantly greater than the FSA—employee value at the 0.05 level

Pearson correlation coefficients:
- All Issues—FCAS(Emp.)/FCAS(Cons.) = 0.9450
- Common issues—FCAS(Emp.)/FSA(Emp.) = 0.8187
- FCAS(Cons.)/FSA(Cons.) = 0.8579

Spearman correlation coefficients:
- All Issues—FCAS(Emp.)/FCAS(Cons.) = 0.9372
- Common issues—FCAS(Emp.)/FSA(Emp.) = 0.8552
- FCAS(Cons.)/FSA(Cons.) = 0.8579
bers of the CAS and their level within their company (senior manager, middle manager, or nonmanagement).

The second survey deals with potential helps and challenges in responding to ethical dilemmas. The survey lists the 16 potential helps and 18 potential challenges found in Tables 2 and 3, respectively. Again, respondents are asked to rate these on a five point scale. For the helps, a 5.0 means the factor is extremely helpful and a 1.0 means it is not helpful. For the challenges, a 5.0 means it presents a significant challenge and a 1.0 means it does not present a challenge. Respondents also are permitted to indicate NA if the helps or challenges are not available or not applicable. Two open-ended questions ask the respondents to indicate other factors they find helpful in resolving ethical dilemmas and other factors that present ethical challenges. Finally, the survey collects demographic information on the number of years the respondent had been a member of the CAS, how long he or she has worked for the current employer, and his or her level within the company.

Both of these surveys are nearly identical to the surveys of FSAs conducted by Cooper and Frank in February 1991. (Minor changes are made to incorporate issues, helps, and challenges that are likely to be relevant for casualty actuaries.) This makes it possible to compare the FSA and FCAS group responses and draw conclusions about their similarities and differences.

2.2 The Samples

The Casualty Actuarial Society supplied two sets of mailing labels for actuaries practicing in the U.S.: 504 property/liability company employees and 248 consultants. It is possible that the ethical dilemmas viewed as important may differ between actuaries employed at insurance companies (which would tend to be larger companies) and those consulting (primarily to smaller companies); therefore, the issues survey was mailed to both company employees and consultants. The helps and challenges survey was mailed only to a sample of company employees.²

²Sending the helps and challenges survey only to company employees while sending the issues survey to both company employees and consultants was done for several reasons. First, given the limited number of FCAS consultants available to survey, the design focused on identifying differences in perceptions of issues by employees and consultants, which both intuition and previous ethics research suggest are more likely to differ for the two groups than are helps and challenges. Second, the design replicates the design of the earlier FSA studies (which reflects discussion with the staff of the SOA) and thus permits comparison between the two groups of actuaries in this paper. The FCAS study split the company employee group—half received the issues survey, and half received the helps and challenges survey. The consultant group only received...
Survey forms were mailed in July 1993. Responses were returned by 84 of the 252 company employees receiving the issues survey (33 percent) and 76 of the 248 consultants (31 percent). The helps and challenges survey generated responses from 112 of the 252 company employees that received it (44 percent).³

3 Findings

3.1 FCAS Groups: Perception of Issues

Table 1 provides the mean ratings for each of the 34 ethical issues based on the individual ratings given to each issue by all respondents, with separate results reported for employees and consultants. The table also shows the rank of each issue based on the size of the issue's mean rating relative to the size of the other issue means for the same group of actuaries. For example, Issue 1 (failure to provide products and services of the highest quality in the eyes of the customer) is rated 2.667 on average by the respondents in the FCAS company employee group and has the second highest issue mean rating for that group.

Only one issue received a mean rating of over 3.0—Issue 8, responding to pressure from clients and/or management to change assumptions used in pricing or reserving. This issue is ranked number one by both the company employees and consultants. While the other issues have mean ratings less than 3.0, the percentage of respondents indicating 3.0, 4.0, or 5.0 for an issue suggests that many of the issues present ethical problems for rather substantial percentages of those in the industry.

The FCAS company employee and consultant groups rank the same six issues as the major ones facing the property/liability insurance industry (although they are in slightly different order in each group). The consultants rate an additional issue (Issue 34) as tied for sixth place. These seven issues and their relative rankings for FCAS employees and consultants are:

³ These response rates are somewhat lower than the FSA response rates. In that case, 48 percent of company employees and 46 percent of consultants responded to the issues survey, and 41 percent responded to the helps and challenges survey. While there are no obvious explanations for the differences in response rates, it should be noted that the CAS had just conducted an extensive survey of its membership when the current survey was mailed.
a) Responding to pressure from clients and/or management to change assumptions used in pricing or reserving (1, 1).  

b) Failure to provide products and services of the highest quality in the eyes of the customer (2, 3).  

c) Failure to identify the customer's needs and recommend products and services that meet these needs (3, 2).  

d) Misrepresenting information provided to regulators (4, 4).  

e) Lack of knowledge or skills to competently perform one's duties (4, 6).  

f) Failure to provide prompt, honest responses to customer inquiries and requests (6, 5).  

g) Misrepresenting or concealing limitations in one's abilities to provide services (8, 6).

Issue 27, office/agency closings and layoffs, ranks relatively high for company employees (and not for consultants). Recent restructuring in the property/liability industry and the effects of the soft market evidently are leaving their effects.

Statistically significant differences between the company employee and consultant groups are reported for four of the seven top-ranked issues (Issues 2, 8, 32, 34). For these four issues (which deal with consumer/client problems) the consultant mean ratings are higher than the employee ratings, perhaps reflecting the consultants' greater contact with clients and others outside the firm. The apparent importance of these differences diminishes, however, when one examines the correlation coefficient for the mean ratings of the 34 issues, 0.9450. This high correlation indicates that the order of the issues is similar for the two groups.

4 For example, (2, 3) means FCAS employees rank the issue second, while FCAS consultants rank the same issue third.

5 t-tests are used to test for differences in means throughout the paper. F statistics are calculated to test for equality in variances. In most cases the variances are not found to be different, so pooled variance t-tests are used in these cases. Where differences in the variances are found, a separate variance t-test is used. The possibility of response bias exists because of the somewhat small response rates (not atypical of this type of study).

6 The numbers reported in the text of this paper are the Pearson product moment correlation coefficients. Because the tables also include ranks, the Spearman rank order correlation coefficients also are calculated and reported in the tables.

7 Little difference is found among actuaries at different organization levels with respect to their order of the issues. All groups rank Issue 8 first. The correlation coefficient for the mean ratings is 0.9354 for senior manager and middle manager company employees and 0.9037 for senior manager and middle manager consultants. The order of issues is somewhat less similar, however, for nonmanagerial personnel and managers for the consultant survey.
As expected, the actuaries rate those issues most closely related to selecting assumptions in pricing and reserving and dealing with regulators among the highest (Issues 8 and 9).\(^8\) Evidence of the importance with which actuaries view these issues is found in the responses to the open-ended question asking for the most important ethical problem facing those working in the industry. Several responses deal with the current regulatory environment, which is considered, as one respondent termed it, "too political." Respondents raised concerns about dealing with consumer advocates and regulators who do not follow professional standards of practice. One respondent stated "Unethical rate suppression leads to unethical support for rate filings."

The remaining five issues identified by the respondents as presenting the greatest ethical problems to the industry combine business and professional ethics issues. Two are related to the ethical responsibilities of businesses and their employees in general (Issues 1 and 2). The other three are ethical issues of special relevance to professionals (Issues 31, 32, and 34). All four of the professional issues included in the survey form (Issues 31 through 34) are ranked in the top ten by both company employees and consultants.

Three of the highest rated issues relate to providing proper service to customers (Issues 1, 2, and 32). While these may be viewed as ethics problems, they are also part of a much broader question of just how the industry does business. Much has been written about the industry's poor relations with consumers and the need to provide quality customer service.\(^9\) Customer service is a major focus of the total quality management (TQM) movement now fashionable with the insurance industry (and others). It is not surprising that these issues are rated highly.

### 3.2 Comparison with FSA Study Findings: Perception of Issues

Table 1 also shows the issue means and rankings for the earlier survey of company employee and consultant FSAs (Cooper and Frank 1992b). FCAS Issues 8 and 9 are not in the FSA survey.\(^{10}\) The correla-

---

\(^8\) For a thorough discussion of the ethical dilemmas faced by actuaries in these areas, see Feldblum (1993).

\(^9\) See, for example, Roberts (1993) who reports on the proposed creation of a permanent quality insurance congress to address the problem of improving the quality of the insurance industry, with particular emphasis on improving service.

\(^{10}\) This is unfortunate because, as one reviewer noted, Issue 8 may be the principal issue for pension actuaries while Issue 9 is common to all actuaries.
The correlation coefficient for the mean ratings of common issues for FCAS employees and FSA employees is 0.8187; for FCAS consultants and FSA consultants it is 0.8579. Not surprisingly (because these actuaries are in essentially different industries), these are lower than the FCAS employee/FCAS consultant correlation coefficient previously reported.

The top issues, as rated by FSAs and FCASs, however, are similar. With the exception of Issues 8 and 9 (which are not included on the FSA survey), all of the top seven FCAS issues also are ranked in the top seven by both the FSA company employees and consultants. The FSAs (both employees and consultants) give the highest rating to Issue 7 (false or misleading representation of products and services in marketing, advertising, or sales efforts), while the FCASs rank Issue 8 highest. The importance of Issue 7 in the life insurance industry likely is driven by a concern over misleading policy illustrations, perhaps the life insurance equivalent to the property/ liability problems with pricing and reserving. Both deal with the fundamental actuarial problem of selecting assumptions to illustrate or predict the future. Also, in both cases the information will be used by outside parties (consumers in the life case, regulators in the property/ liability case).

The issues ranked second and third by the FCASs (Issues 1 and 32) rank fourth and second for the FSAs. The next two highest rated issues (of those common to both studies) for the FCASs (Issues 31 and 2) are also in the top seven FSA issues. These issues (two business-related and two professional-related) reflect a high degree of concern about relations with customers and the industry’s willingness/ ability to provide adequate and appropriate products and service.

3.3 FCAS Survey: Helps

Table 2 provides the mean ratings and ranks for the survey’s 16 potential helps and the percentage of respondents that indicate each help is not available or applicable. (NA responses are not included in the mean ratings.) The potential helps in the survey form are clas-
Vaughan, Cooper, and Frank

Ethical Issues

sified into three groups: one’s personal attributes/environment (H15 and H16), business environment (H5 through H14), and professional environment (H1 through H4). The results indicate that respondents find a number of these resources helpful in resolving ethical dilemmas encountered in work. Only four of the 16 potential helps have mean ratings below 3.0 (H2, H4, H6, and H7).

The five factors ranked most helpful to the respondents for resolving ethical dilemmas encountered in their work are (from highest to lowest):

a) Personal moral values and standards.
b) The fact that your immediate boss does not pressure you into compromising your ethical standards.
c) Ability to go to your boss for information and advice on ethical issues.
d) A company environment/culture that does not encourage you to compromise your ethical values to achieve organizational goals.
e) A company management philosophy that emphasizes ethics in business operations.

Respondents rate their own personal moral values and standards most helpful. All of the remaining four major helps relate to the actuary’s work environment. The actuary’s immediate boss is a major source of help (second and third), and the company culture and management philosophy are close behind (fourth and fifth). Two of these most helpful factors (H10 and H13) deal with merely the absence of pressure to compromise one’s own ethical standards. This suggests that one way companies and managers can assist employees is by neither explicitly nor implicitly pressuring them to go against their ethical values.

In spite of the fact that the company environment appears to be a major source of assistance in resolving ethical dilemmas, many respondents report formal company systems are not available to them. Thirty-eight percent of respondents report that a program or department in their company to which they could report unethical activity (H7) is either not available or not applicable, and 49 percent report company ethics training (H6) as unavailable. Even when these resources are available, respondents tend to rate them among the least helpful factors; both have mean ratings below 3.0. Relatively more helpful is a company code of ethics, which earns a mean rating of 3.4 but is not available to 14 percent of respondents.
<table>
<thead>
<tr>
<th></th>
<th>Potential Helps—All Respondents</th>
<th>FCASs</th>
<th>FSAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>The codes of professional conduct of the Casualty Actuarial Society [Society of Actuaries] and the American Academy of Actuaries</td>
<td>3.327</td>
<td>9</td>
</tr>
<tr>
<td>H2</td>
<td>Materials on ethics published by the Casualty Actuarial Society [Society of Actuaries] and the Academy of Actuaries</td>
<td>2.954</td>
<td>13</td>
</tr>
<tr>
<td>H3</td>
<td>Professional meetings where ethical issues can be discussed with other actuaries</td>
<td>3.140</td>
<td>12</td>
</tr>
<tr>
<td>H4</td>
<td>The Actuarial Board for Counseling and Discipline (ABCD) of the American Academy of Actuaries</td>
<td>2.510</td>
<td>15</td>
</tr>
<tr>
<td>H5</td>
<td>Your company’s code of ethics or similar policy on ethical conduct</td>
<td>3.438</td>
<td>8</td>
</tr>
<tr>
<td>H6</td>
<td>Ethics training provided by your company</td>
<td>2.526</td>
<td>14</td>
</tr>
<tr>
<td>H7</td>
<td>A program or department in your company to which you can report unethical activity</td>
<td>2.300</td>
<td>16</td>
</tr>
<tr>
<td>H8</td>
<td>A company management philosophy that emphasizes ethics in business operations</td>
<td>3.822</td>
<td>5</td>
</tr>
<tr>
<td>H9</td>
<td>Clear communication of appropriate ethical behavior by company management</td>
<td>3.562</td>
<td>7</td>
</tr>
</tbody>
</table>
### TABLE 2 (continued)
**Potential Helps—All Respondents**

<table>
<thead>
<tr>
<th></th>
<th>FCASs Mean Rating</th>
<th>FCASs Rank</th>
<th>FCASs % NA</th>
<th>FSAs Mean Rating</th>
<th>FSAs Rank</th>
<th>FSAs % NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>H10</td>
<td>A company environment/culture that does not encourage you to compromise your ethical values to achieve organizational goals</td>
<td>3.981</td>
<td>4</td>
<td>4</td>
<td>3.91</td>
<td>3</td>
</tr>
<tr>
<td>H11</td>
<td>Ability to go to your boss for information and advice on ethical issues</td>
<td>4.055*</td>
<td>3</td>
<td>3</td>
<td>3.79</td>
<td>4</td>
</tr>
<tr>
<td>H12</td>
<td>Ability to go beyond your boss to higher level managers for information and advice on ethical issues</td>
<td>3.258</td>
<td>10</td>
<td>13</td>
<td>3.21</td>
<td>11</td>
</tr>
<tr>
<td>H13</td>
<td>The fact that your immediate boss does not pressure you into compromising your ethical standards</td>
<td>4.402*</td>
<td>2</td>
<td>4</td>
<td>4.16</td>
<td>2</td>
</tr>
<tr>
<td>H14</td>
<td>Help from your co-workers in resolving your ethical dilemmas</td>
<td>3.248</td>
<td>11</td>
<td>3</td>
<td>3.23</td>
<td>10</td>
</tr>
<tr>
<td>H15</td>
<td>Your own personal moral values and standards</td>
<td>4.795</td>
<td>1</td>
<td>0</td>
<td>4.80</td>
<td>1</td>
</tr>
<tr>
<td>H16</td>
<td>Your family and friends who provide support and insight for you in resolving ethical conflicts</td>
<td>3.706</td>
<td>6</td>
<td>3</td>
<td>3.75</td>
<td>6</td>
</tr>
</tbody>
</table>

* = significantly greater than the FSA value at the 0.05 level  
[ ] = wording on the FSA survey form for these statements  
Pearson correlation coefficient: FCAS/FSA = 0.9838  
Spearman correlation coefficient: FCAS/FSA = 0.9736
Respondents tend to give relatively low ratings to resources provided by the professional societies. Two of the four helps related to professional resources (H1 to H4) receive ratings below 3.0 (H2 and H4). The Actuarial Board for Counseling and Discipline ranks 15th out of the 16 helps, possibly because it is relatively new and respondents have had little experience with it.\(^{13}\) Additional surveys are necessary to determine more precisely the reason for ABCD's relatively low rank.

To summarize, FCAS company employees tend to look first to personal values, second to certain factors in their business environment, and last to resources in their profession when resolving ethical dilemmas. Resources from their profession tend to be viewed as reasonably helpful, but less helpful than most factors in the business environment. The responses also suggest that businesses should ensure that managers are equipped to deal with their own ethical dilemmas and those encountered by their subordinates in the course of work. One of the best things a business can do to encourage ethical behavior is to refrain from pressuring managers and employees to compromise their own personal values.

### 3.4 Comparison with FSA Study Findings: Helps

Table 2 also shows the helps ratings and ranks for the earlier survey of FSAs (Cooper and Frank 1992a). The five most important helps, as rated by respondents to the FSA survey, are identical to those found in the FCAS survey. Moreover, they are in virtually the same order of importance (the exception being H10 and H11, which are reversed in order). The correlation coefficient for mean ratings of the FCAS and FSA respondents is 0.9838, indicating a high degree of similarity in the ordering of their responses. The mean ratings of only two helps (H11 and H13, both of which relate to the respondent's relationship with his or her immediate boss) are significantly different, with the FCAS ratings higher than the FSA ratings.

### 3.5 FCAS Survey: Challenges

Table 3 provides mean ratings by the FCAS company employees for each of the survey's 18 potential challenges and the percentages that report NA. (NA responses are not included when calculating the mean ratings.) All of the mean ratings are below 3.0, indicating that

\(^{13}\) Following the comparison of the FSA and FCAS results, some suggestions for ways the professional societies can improve the helpfulness of professional factors are given.
### TABLE 3
**Potential Challenges—All Respondents**

<table>
<thead>
<tr>
<th>Challenge</th>
<th>FCASs</th>
<th>Mean Rating</th>
<th>Rank</th>
<th>%5</th>
<th>%NA</th>
<th>Rating Rank</th>
<th>FSAs</th>
<th>Mean Rating</th>
<th>Rank</th>
<th>%5</th>
<th>%NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Conflict between duty to your employer and duty to your clients</td>
<td></td>
<td>2.120</td>
<td>4</td>
<td>4</td>
<td>55</td>
<td>2.37</td>
<td>3</td>
<td>5</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2 Conflict between duty to your employer or clients and duty to the</td>
<td></td>
<td>2.660</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regulators or public</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3 Unethical behavior or demands by co-workers</td>
<td></td>
<td>2.019</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>1.98</td>
<td>11</td>
<td>2</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4 Unethical demands made by your employer or clients [by your clients</td>
<td></td>
<td>2.018</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>2.24</td>
<td>5</td>
<td>4</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>or customers]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5 Competition encountered in your business activities</td>
<td></td>
<td>2.708</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>2.90</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6 Intense competition in the insurance industry that forces owners,</td>
<td></td>
<td>2.820</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>3.10</td>
<td>1</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>managers, and others to focus on the bottom line and not business ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7 A company environment/culture that encourages you to compromise your</td>
<td></td>
<td>1.774</td>
<td>14</td>
<td>5</td>
<td>5</td>
<td>1.81</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethical values to achieve organizational goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8 A lack of ethics policy in your company</td>
<td></td>
<td>1.648</td>
<td>17</td>
<td>0</td>
<td>20</td>
<td>1.79</td>
<td>16</td>
<td>3</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9 A lack of communication of ethics policy by your company</td>
<td></td>
<td>1.809</td>
<td>13</td>
<td>2</td>
<td>20</td>
<td>1.90</td>
<td>12</td>
<td>2</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10 A lack of ethics training by your company</td>
<td></td>
<td>1.851</td>
<td>12</td>
<td>1</td>
<td>9</td>
<td>1.98</td>
<td>10</td>
<td>2</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3 (continued)
Potential Challenges—All Respondents

<table>
<thead>
<tr>
<th></th>
<th>Mean Rating</th>
<th>FCASs</th>
<th></th>
<th>Mean Rating</th>
<th>FSAs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank</td>
<td></td>
<td>% 5</td>
<td>% NA</td>
<td>Rank</td>
</tr>
<tr>
<td>C11</td>
<td>Your inability to disclose unethical activity because of fear of management reprisal</td>
<td>1.755</td>
<td>15</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>C12</td>
<td>Mid-level managers who are concerned only with their own personal gain and not ethics</td>
<td>1.907</td>
<td>11</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C13</td>
<td>Measuring employee performance on the basis of end results without also considering how ethical the means were to achieve the results</td>
<td>2.029</td>
<td>6</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>C14</td>
<td>Performance based on quotas such as amounts of insurance sold, cases underwritten, claims processed, or hours billed</td>
<td>2.012</td>
<td>10</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>C15</td>
<td>Pressure from others compensated by commissions [compensation that includes commissions]</td>
<td>2.047</td>
<td>5</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>C16</td>
<td>Compensation that includes bonuses or opportunities for profit sharing</td>
<td>2.020</td>
<td>7</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>C17</td>
<td>A personal need to achieve or succeed</td>
<td>1.746</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C18</td>
<td>Financial pressures exerted on you to provide security for family or others</td>
<td>1.585</td>
<td>18</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

* = significantly greater than the FCAS value at the 0.05 level
[] = wording on the FSA survey form for these statements

Pearson correlation coefficient: FCAS/FSA = 0.9819
Spearman correlation coefficient: FCAS/FSA = 0.9044
the factors generally are not viewed as presenting particularly significant challenges to large percentages of the respondents. Nearly all of the challenges, however, receive a rating of 5.0 (presents a very significant challenge) by at least some of the respondents, and most are rated 3.0, 4.0, or 5.0 by at least 25 percent of respondents. This suggests that these challenges, while not viewed as widespread problems, are sufficiently pervasive that they should not be ignored by management. Managers and supervisors need to be alert to identify and handle (on an individual basis) those situations where reasonably significant challenges are encountered by members of their staffs in attempting to respond appropriately to ethical dilemmas at work.

The five challenges ranked as the most problematic in the respondents' personal efforts to act ethically are (from most to least):

a) Intense competition in the insurance industry that forces owners, managers, and others to focus on the bottom line and not on business ethics.

b) Competition encountered in business activities.

c) Conflict between duty to the employer or clients and duty to regulators or the public.

d) Conflict between duty to the employer and duty to the clients.

e) Pressure from others compensated by commissions.

The two challenges rated the highest deal primarily with competitive pressures. The importance of these pressures as a challenge to behaving ethically is found in previous studies of insurance professionals (e.g., Cooper and Frank 1991a, 1992a). While economic theory suggests that competition is good for business and its owners, perhaps its impact from an ethical standpoint is not always favorable, especially when the rights and obligations of other stakeholders (such as customers, employees, and the general public) are taken into consideration.

That actuaries may have difficulties ascertaining and balancing the rights and obligations of the various stakeholders is seen in the third and fourth highest-rated challenges. Both deal with conflict between duty to one's employer and duty to other stakeholders (public/regulators and clients). The third highest rated challenge (involving conflict between duty to employer and public/regulators) is related closely to Issues 8 and 9 on the issues survey, which also receive high ratings.
3.6 Comparison with FSA Study Findings: Challenges

The factors cited by FCAS respondents as presenting the greatest challenges to their personal efforts to act ethically are similar to those cited by the FSAs in the Cooper and Frank (1992a) study. Three of the top five FCAS challenges (C1, C5, and C6) are also in the top five FSA challenges. One of the remaining two (C4) is not included in the FSA survey, and one (C15) is ranked sixth by the FSAs. Both groups of actuaries rate Challenge 6 (intense competition forcing a focus on the bottom line) as the most troublesome. As in the helps case, the correlation between the FCAS and FSA groups is high, 0.9819.

4 Summary and Conclusions

The ethics literature recognizes that professionals working in a business environment may look to a variety of factors to assist them in resolving ethical dilemmas and that the relative importance placed on these factors may vary across professional groups. Raelin (1989) distinguishes the “cosmopolitan” professional, who “pledges loyalty to the profession” and the “local” professional who extends loyalty to the business organization. A cosmopolitan professional should tend to value professional resources higher than those related to the business environment when resolving ethical dilemmas.

The results of this study indicate that actuaries tend to look first to their own personal values, next to certain factors in their business environment, and only then to professional factors. That is, actuaries tend to be more local than cosmopolitan in their search for solutions to ethical problems. This is not to say that professional factors are unimportant; rather, they tend to be viewed as less helpful than business factors. These results are consistent with prior studies of both insurance and accounting professionals.14

These findings suggest that one way the professional actuarial societies could assist their members is to improve the helpfulness of the professional factors (e.g., the codes of professional conduct, published materials on ethics, professional meetings as a place to discuss ethical issues, and the recently established ABCD). A recent paper by Feldblum (1993) makes a similar suggestion. After examining the ability of the AAA Code of Professional Conduct to distinguish ethical and unethical behaviors, Feldblum recommends that the Code be

14 The relative importance of business factors over professional factors has been found in studies by Cooper and Frank (1991a and b) and Heaston, Cooper, and Frank (1993).
supplemented with guidelines, explanations and interpretations, and case studies.\textsuperscript{15} Some actuaries fear, however, that too much specificity would reduce actuarial work to a cookie cutter approach to solving problems. While there are legitimate differences of opinion over what form (e.g., guidelines, case studies) supplemental materials should take and how rigid and enforceable they should be, it seems clear that the professional societies can do more to improve the helpfulness of professional factors.

The professional societies should recognize in developing a plan for ethical guidance, however, that the business environment tends to be among the first places actuaries look for help and that the boss is among the first places they go within the business environment. Because many actuaries are managers ("bosses"), they have both an opportunity and a responsibility to influence the ethical environment in which they work. The most effective efforts by the professional societies to foster ethical behavior may be those efforts aimed at enhancing their members' ability to contribute to an ethical business environment.

There are two directions that could be taken by the professional organizations to assist their members in influencing the business environment. The first is aimed at changing the \textit{internal} business environment; that is, assisting members to contribute to a corporate culture that encourages and rewards ethical behavior. This would involve, for example, training manager-actuaries to identify and counsel subordinates facing ethical dilemmas. It could involve assisting manager-actuaries in their efforts to introduce ethical resources in their own firms. For example, the professional societies could provide education on how to create an effective corporate code of ethics and how to implement corporate programs providing assistance to employees in resolving ethical dilemmas.

The second direction is aimed at influencing the \textit{external} business environment by addressing the issues that actuaries find most troublesome. For life actuaries, the number one issue is policy illustrations; for casualty actuaries, it is pressure to change assumptions in pricing and reserving. These issues are within the actuary's realm of expertise, but competitive pressures within the industry make it difficult for one individual to act alone. For example, many companies apparently discount loss reserves implicitly on statutory financial state-

\textsuperscript{15} Feldblum concludes that the \textit{Code} is merely a first step to the development of a guide for ethical behavior. "To be universally applicable, the \textit{Code} must be general: it prohibits that which is clearly wrong, but it leaves the ambiguous untouched. In practice, ethical dilemmas come in shades of gray, for which noble precepts provide insufficient guidance."
ments, in spite of statutory prohibitions against discounting.\textsuperscript{16} To the extent this is an industry-wide phenomenon and individual companies are targeted for supervision by the degree to which their financial results vary from the norm, it would be difficult for a single company to behave differently. Professional associations could influence this situation by aggressively pursuing realistic statutory accounting standards. On the life side, the associations could pursue clearly defined guidelines for policy illustrations.\textsuperscript{17}

Both life and casualty actuaries also give high rankings to issues focusing on relations with consumers and providing quality products and services that meet consumer needs. These issues rest on the heart of the major problem facing the industry—the public's lack of trust and the need to rebuild that trust. This is a problem not easily addressed, but one that has received significant attention in recent years. Actuaries should play a major role in rebuilding that trust, both individually (by encouraging an emphasis on ethical and sound business practices) and as a group (by developing industry-wide responses to the problem).

References


\textsuperscript{16} Feldblum (1993) provides a concise description of the ethical dilemma faced by casualty actuaries expressing loss reserve opinions. "The Code is clear: the actuary must set full value reserves if these are required by law or regulation. Insurance practice, which is constrained by business reality, not by the AAA Code, is also clear: many insurers have been setting implicitly discounted reserves, and actuaries have been giving unqualified opinions approving them."

\textsuperscript{17} An example of the part a professional association can play in affecting the external environment is found in recent SOA activity on illustrations. The SOA Task Force for Research on Life Insurance Sales Illustrations recently looked at this problem and recommended that the use of sales illustrations be limited to showing the consumer "the mechanics of the policy being purchased" (Type A usage). The task force concludes that the use of sales illustrations for comparing cost or performance of different policies (Type B usage) is "fundamentally inappropriate" because they are inherently incapable of accounting for differences in risk across policies. While the task force made recommendations for the development of specific standards, disclosures, and regulations for both Type A and Type B usage, it also recommended a long-term strategy of educating users (including agents, home office employees, and consumers) about the appropriate use of illustrations. See \textit{Final Report of the Task Force for Research on Life Insurance Sales Illustrations} (1993). The National Association of Insurance Commissioners has announced its intent to examine the problem of policy illustrations. Among the issues it will address is the appropriateness of Type A and Type B usage. This presents a good opportunity for the AAA to participate in this debate, as suggested by the task force report.


Therese M. Vaughan, Robert W. Cooper, and Garry L. Frank
Drake University
212 Aliber Hall
Des Moines, Iowa 50311
Managing the Relative Volumes of Participating and Nonparticipating Business in a Mutual Life Company

Robert G. Chadburn*

Abstract**

Management decisions of a mutual life company involving the amounts and relative proportions of participating (with profits) and nonparticipating (without profits) business and the level of expenses are examined in relation to their effect on participating policyholders' returns. A particular expense ratio is defined that plays a key role in a framework for making such decisions. The sensitivity of participating policy returns to changes in each factor are analyzed. Companies with expense ratios (as defined) of less than 2 are shown to prefer a different strategy from companies with higher ratios. There is an incomplete tendency for the ratio to stabilize either at unity or to tend to infinity. The practical implications and limitations of the approach are considered.

Key words: decision making; expenses; new business

1 Introduction

This paper concerns certain management decisions relating to mutual life companies (offices); the position regarding stock (proprietary) companies is different and is only briefly discussed.

A United Kingdom (U.K.) environment is assumed, although the circumstances are general enough to make the conclusions appropriate to other countries, including the United States. Some of the comments made and procedures adopted in the paper, however, reflect peculiarities of the U.K. (including methods of dividend distribution,

*Robert G. Chadburn has been a lecturer in actuarial science at the City University, London, since 1989. He has a Ph.D. in population genetics from Liverpool University (1986) and obtained his F.I.A. in 1990. His practical actuarial experience was with the NFU Mutual (1982-1989). Chadburn's current research interests are in life insurance, especially solvency and profit distribution.

** The author is grateful to Miss J. Petty for typing the manuscript and to the editor and three anonymous reviewers for their helpful comments. The author is, however, responsible for all errors that may remain in the paper.
product design, and statutory regulation). Brief descriptions of these features will be given to assist non-U.K. readers.

In the U.K., participating (with profits) policyholders' dividends are paid in two forms, referred to as reversionary and terminal bonuses. Reversionary bonuses are additions to the contractual policy benefit; they usually are made annually, at the discretion of the company's actuary, to reflect a proportion of the surplus earned during the previous year. Terminal bonuses are added at the claim date of the policy, again at the discretion of the actuary, so that the total policy benefit on maturity of a policy will be equal to the policy's asset share plus an element of smoothing. In a mutual company the return to the participating policyholder also will include a share in the company's profits or losses from other sources, such as those generated by nonparticipating business, plus any contribution made to or from the estate.

The nature of the statutory regulations regarding the valuation of assets and liabilities combined with the particular features of the participating business described above result in different patterns of emergence of statutory surplus. Nonparticipating (without profit) business generally produces large initial surplus strains, followed by small regular profits emerging in subsequent years. The large strains, however, can be reduced by modern product designs. Participating business, for which reserves only are required for the contractual benefit plus declared bonuses, lead to reduced or even nonzero initial strains, followed by relatively large contributions to statutory surplus for a considerable period of the policy's duration. A large strain then is produced at the claim date when the terminal bonus becomes payable. As a result, the issue of new participating business will tend to improve the statutory surplus position, while the issue of nonparticipating business will tend to have the opposite effect. This is a factor that will bear on later discussion.

In the U.K., traditional nonparticipating business such as term and whole life insurances do not constitute much of a mutual company's portfolio. A considerable and possibly increasing volume of business consists of unit-linked contracts. In a unit-linked insurance contract, premiums (after deductions for expense and claim charges) are allocated to units, the value of which directly reflect the returns obtained from a specific pool of assets. The charges represent the nonparticipating premium to the company for these contracts.
The decisions considered in this paper are those that ultimately have an effect upon the volumes of new participating and nonparticipating business issued by a mutual company and in the management of expense levels.

According to a basic principle of economics, the more units of product that are sold at the same price for a fixed level of expense, the greater will be the profit per unit sold. Furthermore, an increase in expense levels if accompanied by a greater proportionate increase in units sold will increase unit-profit. This is referred to as *economies of scale*.

In the life insurance business, units of product (policies) are sold, at least partly, with the aim of making a profit and with the knowledge that the activities of selling and managing the business involve expenses that offset profit. A stock company issuing nonparticipating policies will conform ultimately to the basic economic principles stated above, as will a nonparticipating portfolio within a mutual company.

A mutual company, which must have a significant portfolio of participating policyholders on its books, is in an unusual position. As a mutual, all profits earned by both the participating and nonparticipating portfolios must be distributed (ultimately) to the participating policyholders. This means that while increasing the number of participating units sold for a given level of expense will reduce the average cost for each unit sold (thereby increasing unit profit), it also will reduce each unit's share of the profits earned by the nonparticipating portfolio (thereby decreasing unit profit). The position of the mutual company is therefore more complex than the position of a nonparticipating stock company case. The overall profitability of a mutual company depends on the relative levels of profit from the nonparticipating portfolio compared with the level of expenses. It is this position that will be explored in section 3 of this paper.

Profit is not the only consideration of importance to management when arriving at decisions that may affect business volume. For example, the mutual company at all times must maintain a sufficient statutory surplus both to satisfy the regulators and to make investments that are in the best long-term interests of the policyholders, including investment in the issue of new nonparticipating contracts. This surplus is provided by the existence of a participating portfolio, as well as from profits retained from earlier generations of policyholders. A certain relative level of participating business is necessary; without it, a mutual company could not exist.

There are also factors at work in the market that may affect business volume irrespective of any other ambitions the management
may have. For example, sales of nonparticipating contracts may be affected by premium rate, while sales of participating policies may be influenced by historical and current profitability. Customer preferences for products may change over time, and changes to tax legislation (e.g., removal of tax reliefs on insurance premiums) dramatically can influence sales volume. These factors must be borne in mind when considering the implications of the results described in this paper.

The present analysis will need to distinguish between two types of expenses: proportionate and nonproportionate expenses.²

a) *Proportionate expenses* are variable expenses associated with participating and nonparticipating portfolios, and these expenses are proportionate to the volumes of business sold.

b) *Nonproportionate expenses* are the remaining expenses, consisting of other variable expenses and fixed expenses. Nonproportionate expenses can be considered as expenses that collectively vary with the decision made, but not necessarily in proportion to any change in volume of business resulting from the decision.

For example, a particular management decision may lead to an increase in nonproportionate expenses of X percent, coupled with an increase in nonparticipating sales of Y percent; X and Y are not linked to each other in any way other than that they are both dependent upon the decision made. A mutual company attempting to expand its operations to produce economies of scale may be faced with such a decision set. As will be seen in section 3 below, it is always best to choose the decision that produces the greatest increase in sales for the smallest increase in nonproportionate expenses, everything else being equal.

### 2 Construction of Total Profit

All references to present values refer to a time origin (time 0) unless otherwise stated. For the sake of simplicity, it is assumed that the discount rate used to calculate present values is equal to the rate of investment return earned over the lifetime of the portfolio. Further, without loss of generality, it is assumed that the mutual

---

² Chalke (1991) considers expenses at any decision point to be "nonmarginal" if they are invariable by any of the possible decisions made. Expenses that vary according to the decision made are described as "marginal expenses." Ramsay (1991), in his comment on Chalke's paper, points out that these expenses more appropriately are described as "fixed" and "variable" respectively, in accordance with more traditional parlance. Chalke notes that fixed expenses at one decision point may become the variable expenses of the next decision point.
company's business consists of one tranche of nonparticipating business and one tranche of participating business, all issued at time 0. The policies within each tranche are assumed to be identical. The company is assumed to incur three distinct types of expenses:

a) Proportionate expenses of the nonparticipating business;
b) Proportionate expenses of the participating business;
c) Nonproportionate expenses.

The management also has ultimate control of business volume, separately for each tranche.

Three types of profit, $P_n$, $P_{no}$, and $P_w$, need to be defined.

$$P_n = \text{Actuarial present value of future marginal profits (net of proportionate expenses) earned by a single nonparticipating policy issued at time } 0;$$

$$P_{no} = \text{Actuarial present value of future marginal profits (net of proportionate expenses) earned by a single participating policy issued at time } 0; \text{ and}$$

$$P_w = \text{Actuarial present value of the marginal profits earned by a single participating policy including the value of the benefit payments.}$$

Appendix 1 contains a detailed description of the method used to calculate $P_n$ and $P_w$.

While in reality individual policies, even of the same size and type, earn different profits (e.g., due to different dates of claim), it is assumed that each policy earns the same average (or expected) profit. The effect of changes in business volume on profit variability is not considered in this paper.

It is assumed that marginal profits are fixed and independent of sales volume. In practice this is not entirely true: cheaper products are easier to sell, but will have lower marginal profit. In the present context it is helpful to think of the nonparticipating business as a body of unit-linked policies with premium rates that are effectively the charges deducted from the policy benefits. In these cases, policy sales depend more on expected investment returns obtained from the policyholder's unit-holding than upon the rates of charge levied to

---

3 Here *tranche* refers to business issued within a specific and limited time period.

4 Similar conclusions could be drawn assuming the company is in a stationary position, in real terms, issuing constant volumes of new business each year. A single tranche model, however, is much easier to visualize.
cover expenses and other costs, at least up to a point. Hence, an assumption of invariant marginal profit per policy can be justified for the purpose of illustrating the point of interest in this paper. The effect of introducing a price/volume relationship for the nonparticipating business in the model is an aspect worthy of further investigation.

Let

\[ N_n = \text{Number of nonparticipating policies issued at time 0;} \]
\[ N_w = \text{Number of participating policies issued at time 0;} \]
\[ E^{(n)} = \text{Actuarial present value of all future nonproportionate expenses (with respect to these two tranches of business).} \]

The present value of the company's future retained profits from the two tranches, \( TP \), then is given by:

\[ TP = N_n P_n + N_w P_w - E^{(n)}. \]

Because, over the lifetime of the business, all the profits earned by the two tranches are paid to the participating policyholders in policy benefits,\(^5\) it follows that \( TP = 0 \).

Let \( c \) be the present value of future benefits paid to a single participating policy (assumed to be the same for all participating policies), then \( P'_w \) is given by

\[ P'_w = P_w + c. \]

\( P'_w \) can be considered as the value of future premiums, less proportionate expenses, plus the policy's returns on investment. Hence:

\[ 0 = N_n P_n + N_w P'_w - N_w c - E^{(n)} \]

or

\[ c = P'_w + \frac{N_n P_n - E^{(n)}}{N_w}. \]  

\(^5\) This may not always be the case. Smoothing participating policy returns may result in more or less than asset shares being paid, while there may be a strategy to expand or contract the estate for good management reasons. Because policy benefits are designed to follow asset shares and the estate is ultimately a policyholder asset, then it seems appropriate to assume that, on average, \( TP = 0 \).
In other words, the present value of the benefits under a single participating policy is equal to the value of its premiums, including investment income and net of proportionate expenses, plus that policy's share of the profits from the tranche of nonparticipating policies, less that policy's share of the nonproportionate expenses of the company.

From equation (1) it easily can be seen that increasing the volume of nonparticipating business $N_n$, or reducing the amount of nonproportionate expenses $E(n)$, will increase the return to the individual participating policyholder $c$. Increasing the volume of participating business only will increase returns, however, if $(N_nP_n - E(n))$ is negative. That is, the ratio $E(n)/(N_nP_n)$ is greater than unity. This ratio will be referred as $R$, or as the expense ratio,

$$ R = \frac{E(n)}{N_nP_n} $$

and it represents the extent to which the nonproportionate expenses of the portfolio are covered by the nonparticipating business.

The rest of this paper is concerned with identifying the relative effects of varying $N_n$, $N_w$, and $E(n)$ on participating policy returns for different values of $R$. In addition, the paper establishes a framework for the construction of management decisions for companies with particular expense ratios subject to different business prospects.

3 Controlling the Variables to Maintain or Improve Per Policy Profit

3.1 The Variables

It will be assumed that at time 0 management can make decisions that affect $N_n$, $N_w$, $E(n)$, or any combination of these quantities. Equation (2) below represents the value of the participating per policy returns (subsequently referred to as per policy returns) after changes in each of these variables,

$$ (1 + \alpha_c)c = P_w' + \frac{(1 + \alpha_n)N_nP_n - (1 + \alpha'_w)E(n)}{(1 + \alpha_w)N_w} $$

$$ -1 \leq \alpha'_w, \alpha_n < \infty \text{ and } \alpha_w > -1, $$

where $\alpha_n$, $\alpha_w$, $\alpha'_w$, and $\alpha_c$ are parameters indicating the proportional changes in the number of nonparticipating policies, number of
participating policies, nonparticipating expenses, and per policy returns respectively.

**3.2 Maintaining Returns**

Whenever conditions change, it is reasonable to assume that the aim of management will be to ensure that per policy returns do not fall, (i.e., to ensure that \( \alpha_c \) is never negative). Subtracting equation (1) from equation (2) yields:

\[
\alpha_c c = \frac{(1 + \alpha_n)N_nP_n - (1 + \alpha_c^{(n)})E^{(n)}}{(1 + \alpha_w)N_w} - \frac{N_nP_n - E^{(n)}}{N_w}
\]

\[
= \frac{N_nP_n}{N_w} \left[ \frac{(1 + \alpha_n) - (1 + \alpha_c^{(n)})R}{1 + \alpha_w} - (1 - R) \right]
\]

which implies:

\[
\alpha_c = \frac{N_nP_n}{cN_w} \left[ \frac{\alpha_n - (1-R)\alpha_w - R\alpha_c^{(n)}}{1+\alpha_w} \right].
\]  

It is instructive to examine the behavior of \( \alpha_c \) with respect to the other parameters. From equation (3),

\[
\frac{\partial \alpha_c}{\partial \alpha_n} = \frac{1}{(1+\alpha_w)} \frac{N_nP_n}{cN_w} \geq 0
\]  

because the constants \( c \) and \( N_w \) are positive and \( N_n \) and \( P_n \) are non-negative. Notice that the right side of equation (4) is independent of \( \alpha_n, \alpha_c^{(n)}, \) and \( R. \) Thus, returns increase at a constant rate for any change in these quantities.

Similarly,

\[
\frac{\partial \alpha_c}{\partial \alpha_w} = \frac{- (1-R)}{(1+\alpha_w)^2} \frac{N_nP_n}{cN_w}
\]

and

\[
\frac{\partial \alpha_c}{\partial \alpha_c^{(n)}} = \frac{- R}{(1+\alpha_w)} \frac{N_nP_n}{cN_w} \leq 0.
\]
From equation (5), returns (as a function of $\alpha_w$) either are decreasing, zero, or increasing if $R < 1$, $R = 1$, or $R > 1$, respectively. Finally, from equation (6) we see that for a given $\alpha_w$, returns decrease at a constant rate regardless of the level of $\alpha_e^{(n)}$.

Let us now investigate the behavior of $\alpha_n$, $\alpha_w$, and $\alpha_e^{(n)}$ when there is no change in the level of returns; that is, when $\alpha_c = 0$. First, setting $\alpha_c = 0$ in equation (3) yields:

$$\alpha_n = (1-R)\alpha_w + R\alpha_e^{(n)}.$$  

That is, to maintain returns, the proportional change in the number of nonparticipating policies ($\alpha_n$) must be a weighted average of the proportional change in the number of participating policies ($\alpha_w$) and the change in nonproportional expenses ($\alpha_e^{(n)}$). Here the weights can be negative (if $R > 1$). When $0 < R < 1$, in order for returns to be maintained, nonparticipating business has to be increased in response to increases in both expenses and participating business. Decreases in $E^{(n)}$ and $N^{(w)}$ would allow nonparticipating volume to fall while maintaining returns.

Consider the following pairs of parameters: $(\alpha_n, \alpha_w)$, $(\alpha_n, \alpha_e^{(n)})$, and $(\alpha_w, \alpha_e^{(n)})$ in equation (7), subject to the third parameter being set equal to zero. Define $f_{x/y}$ as:

$$f_{x/y} = \frac{\alpha_x}{\alpha_y}$$

where $(\alpha_x, \alpha_y)$ is one of the pairs of parameters listed above and subject to the constraints of equation (7). In other words, $\alpha_x$ is the change in the factor identified by $x$ which is exactly sufficient to maintain returns (i.e., $\alpha_c = 0$) following a change of $\alpha_y$ in the factor identified by $y$ and no change in the third factor in equation (7).

**Definition 1**
When $|f| < 1$, the response is termed efficient; when $|f| \geq 1$, the response is termed inefficient.

**Definition 2**
If $|f_{x/z}| < |f_{y/z}|$, then a change in $z$ is compensated for more efficiently (or less inefficiently) by changing $x$ rather than $y$.

Consider the pair $(\alpha_n, \alpha_w)$. By setting $\alpha_e^{(n)} = 0$ in equation (7), we have $\alpha_n = (1-R)\alpha_w$ which implies that:
R.G. Chadburn  

Participating and Nonparticipating Business in Mutual Life

\[
\frac{\alpha_n}{\alpha_w} = f_{n/w} = (1-R)
\]

Similarly, setting \( \alpha_w = 0 \) gives \( \alpha_n = R\alpha_e^{(n)} \) and

\[
\frac{\alpha_n}{\alpha_e^{(n)}} = f_{n/e} = R
\]

while setting \( \alpha_n = 0 \) gives \((1-R)\alpha_w + R\alpha_e^{(n)} = 0\) and

\[
\frac{\alpha_w}{\alpha_e^{(n)}} = f_{w/e} = \frac{R}{R-1}.
\]

The following results are derived easily from Definition 1:

<table>
<thead>
<tr>
<th>Efficient Region</th>
<th>Inefficient Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_{n/w} = (1-R) )</td>
<td>( 0 &lt; R &lt; 2 )</td>
</tr>
<tr>
<td>( f_{n/e} = R )</td>
<td>( 0 &lt; R &lt; 1 )</td>
</tr>
<tr>
<td>( f_{w/e} = \frac{R}{R-1} )</td>
<td>( 0 &lt; R &lt; 1/2 )</td>
</tr>
</tbody>
</table>

Tables 1 through 3 display summary information on the effects of controlling various parameters to maintain per policy returns.

**TABLE 1**

Summary of the Nonparticipating Business Response With Respect to Changes in Expenses and Volume of Participating Business in Order to Maintain per Policy Returns

<table>
<thead>
<tr>
<th>R</th>
<th>Nonparticipating Response</th>
<th>Due to Nonparticipating:</th>
<th>Due to Expenses:</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1/2)</td>
<td>INC</td>
<td>INC</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>(1/2, 1)</td>
<td>INC</td>
<td>INC</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>(1, 2)</td>
<td>INC</td>
<td>DEC</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>(2, (\infty))</td>
<td>INC</td>
<td>DEC</td>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

INC = Increase  
DEC = Decrease  
E = Efficient  
I = Inefficient
TABLE 2
Summary of the Participating Business Response
With Respect to Changes in Expenses
and Volume of Nonparticipating Business
in Order to Maintain Per Policy Returns

| R       | Participating Due to Due to Notes |
|---------|----------------------------------|-------------------------------|
|         | Response | Nonparticipating: Expenses: |                               |
| (0, 1/2)| DEC      | DEC | I | INC | E | No solution where $R(1+\alpha_n) > (1+\alpha_w)$ |
| (1/2, 1)| DEC      | DEC | I | INC | I |                               |
| (1, 2)  | INC      | DEC | I | INC | I |                               |
| (2, $\infty$) | INC | DEC | E | INC | I |                               |

For key, see bottom of Table 1

TABLE 3
Summary of the Expenses Response With Respect to
Changes in the Volumes of Nonparticipating and Participating Business
in Order to Maintain Per Policy Returns

| R       | Expenses Due to Due to Notes |
|---------|-----------------------------|-------------------------------|
|         | Response | Participating: Nonparticipating: |                               |
| (0, 1/2)| DEC      | INC | I | DEC | I | No solution where $\alpha_n < \alpha_w - R(1+\alpha_w)$ |
| (1/2, 1)| DEC      | INC | E | DEC | I |                               |
| (1, 2)  | DEC      | DEC | E | DEC | E |                               |
| (2, $\infty$) | DEC | DEC | E | DEC | E |                               |

For key, see bottom of Table 1

4 Sensitivity Analysis

The extent to which changes in the three factors affect the per policy returns now will be analyzed using a hypothetical model company.

The model company is composed entirely of 10 year annual premium pure endowments, with one tranche in unit-linked (nonparticipating) form, the other as participating. The methodology used to calculate $P_n$, $P_{nw}$, and $E(n)$ are described fully in Appendix 1. The assumptions used to calculate $P_n$ and $P_{nw}$ are given in Appendix 2. These assumptions lead to $P_n = £255.69$; $P_{nw} = £3517.45$

The present value of future nonproportionate expenses $E(n)$ is calculated such that $P_n$ less one policy’s share of these expenses is equal to 50 percent of the initial commission (IC), i.e.,

$$0.5 \times IC = P_n - \frac{E(n)}{N_n + N_{nw}}.$$

This implies that:
\[ E^{(n)} = (N_n + N_w) (P_n - 0.5 \times IC) \]
\[ = 105.69 (N_n + N_w) \]

(according to these assumptions).

The present value of the participating maturity benefit \( c \) is calculated according to equation (1). The participating policy is assumed to have a sum assured \( S \) such that a compound reversionary bonus of 5 percent per annum (with no terminal bonus) will lead to the implied maturity value of \( c \times (1.1)^{10} \), i.e.,
\[ S = c \times \left( \frac{1.1}{1.05} \right)^{10} . \]

The analysis involves calculating \( c(1+\alpha_c) \) using equation (2), produced for values of \( \alpha_c \) of +0.5 and -0.5, for each of the three factors in turn for \( R = 0.5, 0.75, 1.0, 1.5, 2.0, 3.0 \). Note that \( \alpha_c \) can be expressed in terms of the implied revised reversionary bonus rate \( r \), which satisfies:
\[ S(1+r)^{10} = c \times (1+\alpha_c)(1.1)^{10} . \]

The results are given in Table 4, and the changed values of \( R \) which correspond to these revised bonus rates are given in Table 5.

**TABLE 4**

<table>
<thead>
<tr>
<th>( N_n )</th>
<th>( \alpha_n = 0.5 )</th>
<th>( \alpha_n = -0.5 )</th>
<th>( \alpha_w = 0.5 )</th>
<th>( \alpha_w = -0.5 )</th>
<th>( \alpha^{(n)}_w = 0.5 )</th>
<th>( \alpha^{(n)}_w = -0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>4771.00</td>
<td>5.46</td>
<td>6.33</td>
<td>4.47</td>
<td>6.46</td>
<td>4.20</td>
</tr>
<tr>
<td>0.75</td>
<td>1228.00</td>
<td>5.45</td>
<td>4.53</td>
<td>4.92</td>
<td>5.23</td>
<td>4.65</td>
</tr>
<tr>
<td>1.00</td>
<td>704.62</td>
<td>5.27</td>
<td>4.73</td>
<td>5.00</td>
<td>5.00</td>
<td>4.73</td>
</tr>
<tr>
<td>1.50</td>
<td>380.40</td>
<td>5.15</td>
<td>4.86</td>
<td>5.05</td>
<td>4.86</td>
<td>4.78</td>
</tr>
<tr>
<td>2.00</td>
<td>260.50</td>
<td>5.10</td>
<td>4.90</td>
<td>5.07</td>
<td>4.80</td>
<td>4.80</td>
</tr>
<tr>
<td>3.00</td>
<td>159.80</td>
<td>5.06</td>
<td>4.94</td>
<td>5.08</td>
<td>4.75</td>
<td>4.81</td>
</tr>
</tbody>
</table>
5 Discussion

5.1 Interpretation of the Results

In this section reference will be made particularly to Tables 1 to 4 and to equations (9) to (11).

Consider first Table 4. Sensitivity to changes varies both according to the company’s expense ratio, \( R \), and according to the factor involved. Returns become extremely sensitive at expense ratios below 0.5. But as these values imply high nonparticipating volumes coupled with low expenses, ratios in this region are unlikely in mutual life companies, which need a substantial volume of participating business to be viable.

As a general observation, yield becomes less sensitive to changes the higher is the expense ratio. At values of \( R \) above about 1.5 the improvements in yield due to increasing the volume of either types of business are barely appreciable. For these values of \( R \), the greatest improvements are achieved by reducing nonproportionate expense levels.

At values of \( R \) above about 2, the most significant adverse effect is due to a decrease in the volume of participating business; hence, maintaining the volume of this business should be of most concern to a company with such a ratio. From Definition 2, \(|f_w/e| < |f_n/w|\) indicates that an unavoidable fall in participating volume is much more efficiently dealt with by decreasing expenses than by increasing nonparticipating volume. This difference in efficiency becomes more marked for increasingly large values of \( R \). Similarly, an increase in expenses is compensated for more efficiently by increasing participating rather than nonparticipating volume (\(|f_w/e| < |f_n/e|\)).

Offices with ratios between 1 and 2 should become more concerned with falls in nonparticipating volume and increases in expense levels. Reducing expense levels is a much more efficient way of compensating for a fall in nonparticipating volume than increasing participating volume (\(|f_e/n| < |f_w/n|\)). At ratios close to unity, varying the partic-

<table>
<thead>
<tr>
<th>TABLE 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changes to the Expense Ratio ( R ) After 50 Percent Variations in Fixed Expenses and in the Volumes of Nonparticipating and Participating Business, Where These Values Correspond to the Same Changed Situations That Produce the Yields Shown in Table 4 at any Given Value of ( R )</td>
</tr>
<tr>
<td>( \alpha_n = 0.5 )</td>
</tr>
<tr>
<td>( 0.67R )</td>
</tr>
</tbody>
</table>

39
ipating volume will have almost no effect on yield. There is no efficient way to deal with increasing expenses at these ratios; hence, this would appear to be the most significant problem. If increasing expenses is unavoidable, then increasing the nonparticipating volume is the least inefficient way of compensating ($|f_{nP}| < |f_{wP}|$). The greatest improvements at these ratios can be achieved by increasing nonparticipating volume or by decreasing the nonproportionate expenses.

At ratios below unity a rather peculiar and apparently unstable situation exists, as per policy returns increase with a fall in participating volume, reflecting the increased share of the (positive) value of $(N_nP_n - E^{(n)})$ per participating policy. Returns become increasingly sensitive to changes in all factors, but particularly to changes in the nonparticipating volume. There is no efficient way of compensating for a fall in nonparticipating volume at these levels—it is of particular concern to management to maintain nonparticipating volume here. Between ratios of 0.5 and 1, $|f_{nP}| < |f_{wP}|$, i.e., it is less inefficient to compensate for falling nonparticipating business by reducing expenses than by decreasing participating sales; however, the opposite is the case for the (rather unlikely) situation where the expense ratio is below 0.5.

At ratios below unity, $|f_{nP}| < 1$, so that an increase in expenses can be compensated for efficiently by increasing nonparticipating volume. Reducing the participating business is also an efficient way of dealing with increased expenses at ratios below 0.5, although the nonparticipating route is always the most efficient method.

### 5.2 Consequences of Management Decisions

A company with an expense ratio exceeding 2 would be most concerned with maintaining and increasing participating sales and controlling expenses. Economies of scale are easier to achieve using participating sales the larger the value of the expense ratio. But in all cases, a greater proportionate increase in sales than in expenses is needed to secure these economies. These actions would tend to increase the expense ratio, making it proportionately easier to achieve further economies of scale. The high ratio position persists and tends to become increasingly stable as $R \to \infty$.

At expense ratios in the region $1 < R < 2$, it becomes increasingly easier (in proportionate terms) to maintain or to improve returns by increasing nonparticipating business or by reducing expenses. Economies of scale best would be achieved by increasing nonparticipating sales, although the proportionate increase in sales has to be
larger than that of the expenses. All these actions would result in yet lower expense ratios, making economies of scale easier to achieve and hence continuing the reduction in expense ratio to unity. Any attempt to obtain economies of scale by increasing the participating portfolio becomes increasingly difficult and inefficient, the closer the expense ratio is to 1 from 2. If successful, though, such an action would tend to increase the ratio.

At expense ratios below unity, economies of scale can be achieved efficiently by increasing the nonparticipating business (i.e., if the result of the decision is for \( 1 > \alpha_n > f_{n/e} \times \alpha_k^{(n)} \)). This action (i.e., efficiently producing economies of scale) would tend to increase the expense ratio toward unity. Even greater returns could be achieved if \( \alpha_n > 1 \), in which case the ratio will reduce. Participating sales, however, cannot be increased without lowering per policy profit (or at least without increasing the nonparticipating portfolio sufficiently to compensate for the losses). On the other hand, a company in such a position may be providing higher returns than its market competitors, other things being equal. Such returns would make the company attractive to new participating policyholders, who would accept a fall in per policy profit just to obtain a share of some of it; alternatively, the company could be a potential candidate for demutualization. Hence, market forces could act to increase the participating portfolio—if this ultimately leads to increases in nonproportionate expenses, then this also will increase the expense ratio. Another alternative is for the company to reduce its nonparticipating premium rates (or charges), which would tend to increase the expense ratio as \( P_n \) would be reduced. This effectively transfers some of the superprofits to the nonparticipating policyholders, an action that might be required on the grounds of equity. The need to maintain the participating portfolio in order to provide an adequate statutory surplus also should be borne in mind.

If the only consideration of management is to increase per policy returns, then once \( R < 1 \) the optimum decision would be to reduce the participating portfolio down to one policy. Market forces, coupled with the company's need to provide capital, would tend to reverse the trend. The ultimate position (i.e., value of \( R \)) at which a company would tend to maintain itself would be largely dependent upon the market level of per policy profits expected from participating policies, although there is a partially stable point at \( R = 1 \) caused by attempts to produce economies of scale through efficient increases to the nonparticipating portfolio.
There are two distinct strategies that a company can adopt to maintain a required level of profit, associated respectively with low and high expense ratios.

a) **Low Ratio Strategy**—A company with an expense ratio in the region of unity would be in a highly manageable position. With all nonproportionate expenses covered by nonparticipating business, participating volume can be increased or decreased with no change to returns, provided the statutory solvency position is not compromised by any decrease in volume. Control of per policy profit would rest entirely with controlling the volume of nonparticipating business and level of nonproportionate expenses (and in controlling the expense ratio). Market demand for profit levels would tend to dictate where the ratio ultimately would lie, although pursuit of economies of scale introduces a partial optimum expense ratio at unity itself.

b) **High Ratio Strategy**—A company with a high expense ratio implies that nonparticipating business is essentially an insignificant proportion of the portfolio. Control of per policy profit would rest almost entirely with controlling the volume of participating business and the level of nonproportionate expenses, while pursuit of economies of scale would tend to increase the expense ratio still further.

### 5.3 Practical Implications

The main implications from the above are for mutual life companies that maintain significant volumes of nonparticipating (including unit-linked) business, implying low expense ratios and hence requiring a low ratio strategy. The lower the ratio, the more sensitive per policy profits are to changes in the constituents of the expense ratio. At ratios less than unity, the fact that increasing participating business reduces profit should be borne in mind. At ratios near unity, management should bear in mind that no increase (or decrease) in the participating portfolio will affect returns; a policy of expansion involving increasing expenses matched by increasing the sales of participating contracts would have only adverse effects on returns. Such companies also need to consider the need to meet statutory solvency levels, always an important consideration where significant levels of nonparticipating business are involved.

The actual value of the ratio for any particular company will determine the required response to adopt for any particular situation: for example, when pursuing economies of scale, in determining the minimum increase required to the nonparticipating portfolio to cover an increase in expense levels. Other management decisions that can
be assisted by the response relationships described in this paper include:

a) How can market share be increased most efficiently in order to make minimum losses/maximum profits for the participating policyholders?

b) When business is falling, to what extent should expenses be reduced and which type of business is it most important to retain?

6 Summary

The ratio of nonproportionate expenses to total marginal profits from nonparticipating business (the expense ratio) is a key factor in determining management policy regarding business volume and expenses.

Relationships presented in this paper can be used to determine minimum responses required to compensate for changes in any of the factors in order to maintain per policy returns and also to assist management in choosing appropriate strategies for achieving such aims as economies of scale, increasing market share, or cost-cutting.

Decision choices should vary depending on whether the company has a low expense ratio (less than 2) or a high expense ratio (greater than 2). There are two partially optimum ratios, at $R = 1$ and $R \to \infty$, both resulting from companies choosing the most efficient methods to produce economies of scale at $R < 2$ and $R > 2$, respectively.

It is not sufficient to assume that increasing sales of participating contracts will improve per policy returns, as the greater coverage of expenses is offset to a greater or lesser extent by the dilution of profits from the nonparticipating portfolio.

Particularly at low expense ratios, decision choices identified by the relationships described in this paper will be constrained by the need to meet statutory solvency levels and to maintain adequate investment flexibility. Other factors, such as market forces, also can affect levels of business. All relevant factors should be considered together.

References


Appendix 1

Let

\[ w_{q_x} = \text{Probability of a participating policyholder at age } x \text{ dying before age } x + 1; \]
\[ n_{q_x} = \text{Probability of a nonparticipating policyholder at age } x \text{ dying before age } x + 1; \]
\[ w_{P_x} = 1 - w_{q_x}; \]
\[ n_{P_x} = 1 - n_{q_x}. \]

Assume that all participating policies are \( t \) year pure endowments issued to a life age \( x \). If a participating policyholder dies before the policy matures at age \( x + t \), there is a return of the accumulated fund at the end of the year of death. The fund is set equal to the participating policy’s asset share on death (including its share of nonparticipating policy profits and its share of nonproportionate expenses). Premiums are level and are paid for \( t \) years.

Define \( F_k \) to be the expected fund at time \( k \) immediately before the payment of the death benefit:

\[
F_k = wF_k + nF_k - eF_k \quad \text{(A.1)}
\]

where \( wF_k, nF_k \), and \( eF_k \) are defined below.

\[
wF_k = N_w \times \prod_{r=0}^{k-1} \left( G - wE_r^{(p)} \right)(1 + i)^{k-r} \quad \text{(A.2)}
\]

\[
nF_k = N_n \times \sum_{r=0}^{k-1} w_{P_x+r} \times n_{P_x+r} \times \left( H_r - nE_r^{(p)} \right)(1 + i)^{k-r} \quad \text{(A.3)}
\]

\[
eF_k = \sum_{r=0}^{k-1} w_{P_x+r} \times E_r^{(n)}(1 + i)^{k-r} \quad \text{(A.4)}
\]

where:

\[
v = (1+i)^{-1};
\]
\[
G = \text{Annual gross premium;}
\]
\[
wE_r^{(p)} = \text{Per policy proportionate expenses for a single participating policy paid at time } r;
\]
\[
nE_r^{(p)} = \text{Per policy proportionate expenses for a single nonparticipating policy paid at time } r;
\]
\[
nH_r = \text{Charges paid at time } r \text{ per nonparticipating policy; and}
\]
Total nonproportional expenses at time \( r \).

Note that the payment \((n H_r - n E_r^{(p)})\) at time \( r \) depends on the survival of nonparticipating policyholders to time \( r \). What remains of these payments by time \( k - 1 \) depends on how many participating policyholders survive to time \( k - 1 \). Clearly \( wF_k, nF_k, \) and \( eF_k \) are actuarial "accumulated" values up to time \( k - 1 \) (including both interest and mortality) and from time \( k - 1 \) to time \( k \) using interest only.

The expected actuarial present value of future claims per participant policy is \( c \) where

\[
c = \frac{\sum_{k=0}^{t-1} v^{k+1} \times w q_{x+k} \times F_{k+1} + v^t \times w p_{x+t-1} \times F_t}{N_w}.
\]

(A.5)

Next, define \( P_w', P_n, \) and \( E^{(n)} \) as follows:

\[
P_w' = \sum_{k=0}^{t-1} v^{k+1} \times w q_{x+k} \times \frac{wF_{k+1}}{N_w} + v^t \times w p_{x+t-1} \times \frac{wF_t}{N_w} = \sum_{k=0}^{t-1} w k p_x \times w q_{x+k} \sum_{r=0}^{k} (G - wE^{(p)}) v^r
\]

\[
+ v^t p_x \sum_{r=0}^{t-1} (G - wE^{(p)}) v^r
\]

(A.6)

\[
P_n = \sum_{k=0}^{t-1} v^{k+1} \times w q_{x+k} \times \frac{nF_{k+1}}{N_n} + v^t \times w p_{x+t-1} \times \frac{nF_t}{N_n} = \sum_{k=0}^{t-1} w q_{x+k} \sum_{r=0}^{k} \frac{w}{k-r} p_x \times (H_r - nE^{(p)}) v^r
\]

\[
+ \sum_{r=0}^{t-1} \frac{w}{t-r} p_{x+r} \times (H_r - nE^{(p)}) v^r
\]

(A.7)

and

\[
E^{(n)} = \sum_{k=0}^{t-1} v^{k+1} \times w q_{x+k} \times eF_{k+1}
\]
From the definition of c in equation (A.5),

\[
c = \sum_{k=0}^{t-1} \nu^{k+1} \times w q_{x+k} \left[ \frac{w F_{n+1}}{N_w} + \frac{N_n}{N_w} \times \frac{n F_{k+1}}{N_n} - \frac{e F_{k+1}}{N_w} \right] \\
+ \nu^t \times \frac{w}{n} q_{x+t-1} \left[ \frac{w F_t}{N_w} + \frac{N_n}{N_w} \times \frac{n F_t}{N_n} - \frac{e F_t}{N_w} \right] \\
= P_{n,w} + \frac{N_n P_n - E(n)}{N_w} \\
= P_{n,w} + \frac{N_n P_n - E(n)}{N_w} .
\]
Appendix 2—Model Office Assumptions

Annual premium = £600

Proportionate Expenses

Initial commission = 50 percent of annual premium
Renewal commission = 2.5 percent of annual premium
Other initial expenses = £60
Investment expenses = 0.25 percent of accumulated asset share at end of each year
Other renewal expenses = £6 per annum, inflating at 7.5 percent per annum

Charges for Unit-Linked Policy

Initial = £500
Renewal for commission = 2.5 percent of annual premium
Renewal for fund management charge = 0.5 percent of unit fund at end of each year
Renewal for other = £15 inflating at 7.5 percent per annum

Other Assumptions

Asset accumulation rate = 10 percent per annum
Rates of mortality and withdrawal = nil
Tax rates = nil
Discount rate for calculating present values = 10 percent per annum

R.G. Chadburn
Department of Actuarial Science & Statistics
The City University
Northampton Square
London EC1V 0HB
A Critique of Defined Contribution Plans Using a Simulation Approach

David M. Knox*

Abstract**

During the 1980s there was a trend in many countries away from defined benefit plans toward defined contribution plans. This development means that the individual member bears the full investment risk in the preretirement period and the annuity rate risk at retirement, as no pension benefit (expressed as a percentage of salary) is provided.

This paper, through the use of a stochastic model for both inflation and a range of investment returns, analyses the distribution of retirement incomes that will be produced from a defined contribution plan. The impacts of changing entry and exit ages, different investment strategies, alternative career paths, and different economic assumptions also are assessed. The uncertainty of the resulting income benefits is highlighted, and the question is raised as to whether the individual member is aware of these results.

Key words: funding, pensions, risk

1 Introduction

The provision of retirement income for employees traditionally has been initiated by employers through a defined benefit scheme providing pension benefits. During the last decade, however, there has been a significant shift in many countries toward the provision of retirement benefits through defined contribution plans (or money purchase arrangements). The reasons for this trend vary between countries, but include:

* David Knox is the Foundation Professor at the University of Melbourne and director of the Centre for Actuarial Studies. He previously has taught at Macquarie University and at the University of Waterloo. His recent research interests have concentrated on some of the broader taxation and social policy issues in the superannuation and pensions area.

** I wish to give my sincere thanks to Ms. Ying Teoh who prepared the program necessary to undertake these simulations. She worked with much enthusiasm and did a great job, especially as the number of variables continued to grow! In addition, my thanks to the anonymous referees for their comments that improved this paper.
a) The desire by some employers to reduce their risk present within a defined benefit scheme.
b) Increasing legislation, which often has made defined benefit plans more complex and costly to administer.
c) The presence of surplus in many defined benefit plans and the related issues of overfunding, which may have been encouraged by conservative actuarial assumptions.
d) The high rates of return in the 1980s which made defined contribution plans more attractive to members.
e) The trend toward individual responsibility and the desire by many governments for employees to accept greater responsibility in providing their retirement benefits (for example, with reductions in social security benefits in many countries).
f) The increasing levels of vesting and preservation required by many governments often have been expressed in terms of members' accumulated contributions.
g) Changing taxation structures that permitted and encouraged defined contribution arrangements.

The extent of this trend varies between countries, but it is present in sufficient countries to suggest a significant and long-term direction.

For example, within the United States the number of defined benefit plans decreased 16.7 percent in the five years to 1988 while the number of defined contribution plans increased 36.5 percent (Turner and Beller, 1992). In the same period, the level of contributions to defined benefit private pension plans decreased 43.2 percent to $26.3 billion in 1988 while the level of contributions to defined contribution plans increased 79.7 percent to $64.9 billion in 1988. Turner and Beller (1992, p. 9) note “the gradual but steady replacement of defined benefit plans by defined contribution plans as the primary vehicle for providing pension benefits.” In many, but not all, of these cases the defined contribution benefit represents a benefit in addition to a pension from a defined benefit scheme.

Within the United Kingdom, the trend toward defined contributions plans has not been as strong. The introduction of personal portable pensions in 1988 with the associated legislation, however, has meant that many individuals have been encouraged to contribute to a money purchase (or defined contribution) arrangement.

The recent Australian experience also reflects the move toward defined contribution plans. In 1987 a national industrial agreement was handed down that granted most workers an employer contribution equal to 3 percent of earnings. In July 1992 this approach was
extended so that all employees now receive a minimum employer contribution of either 3 percent or 5 percent of earnings (depending on the size of the company). This minimum employer contribution will increase to 9 percent of earnings by 2002. Although defined benefit plans are permitted and remain with many larger employers, the legislation expresses the minimum contributions in terms of current earnings which represents a defined contribution approach.

This trend toward an increased reliance on defined contribution funds to provide employees' retirement benefits needs to be assessed in terms of the ultimate benefit provided to the member. Actuaries are aware that within a defined benefit pension scheme, the employer bears the investment risk, the salary inflation risk, and the longevity risk (if an annuity is not purchased by the fund). Within a defined contribution plan (where the employer contribution is set as a fixed percentage of the employee's earnings and the final benefit represents the accumulation of these contributions), however, the employer bears none of these risks. Indeed, all risks have been passed to the employee. If employees increasingly are bearing these risks, it is essential that policy makers, individual members, and the pension industry fully understand these risks. With this objective in mind, this research analyses the defined contribution arrangements from the member's perspective.

The paper will consider the benefits that arise from a contribution rate (fixed as a percentage of salary) allowing for stochastic investment and inflation rates and changes in a number of parameters, including contrasting investment strategies, different entry and retirement ages, fractional and full-time employment patterns, and the impact of different annuity rates available at retirement. A fixed 12 percent contribution rate has been chosen, as it provides an adequate retirement pension, on average, for a person who is a member for about 40 years if there are no social security contributions and benefits. For countries with compulsory social security, a lower level of fixed contributions would be appropriate and the benefits can be reduced proportionately. In Australia, where there exists no universal social security benefits, the government has a long-term objective of a total contribution rate equal to 12 percent of earnings.

2 The Model

2.1 Accumulation of Contributions and the Benefits Arising

During an individual's preretirement years, it is assumed that contributions (expressed as a percentage of annual earnings) will be
David M. Knox
A Critique of Defined Contribution Plans

paid mid-way through each year and that investment income will be generated until retirement age. Allowance also can made for any tax payments on contributions and investment income. In many countries (for example, the United States, Canada, and most European countries) contributions and investment income are tax exempt, so the relevant tax variables \( TAX_c \) and \( TAX_I \) can be set to zero without affecting the model. A country where these rates are not zero is Australia where both employer contributions and investment income are taxed at a rate of 15 percent, although the investment tax rate normally is reduced to a net rate between 5 percent and 10 percent due to the availability of various credits.

Equation (1) represents the accumulated contributions available at retirement age for the provision of retirement income. Let \( ACR \) be an employee’s accumulated contributions after \( R \) years in the plan; then:

\[
ACR = K (1-TAX_c) \sum_{t=0}^{R-1} F_t SAL_t (1 + INV_t [1-TAX_I])^{1/2} \\
\times \prod_{u=t+1}^{R-1} (1+INV_u [1-TAX_I])
\]

where:

- \( K \) = Rate of contributions as a percentage of earnings;
- \( TAX_c \) = Rate of tax on the contributions, paid at the time of payment;
- \( TAX_I \) = Net rate of tax on investment earnings;
- \( F_t \) = Fraction of full time employment in year \( t \) (to allow for part timers);
- \( SAL_t \) = Annual salary in year \( t \);
- \( INV_t \) = Gross rate of investment return earned in year \( t \);
- \( R \) = Number of years in the plan before retirement.

For the purposes of this paper, it will be assumed that this accumulated amount will purchase an indexed annuity (or pension) payable for life from the age of retirement. The value of the pension purchased can be expressed as follows:

\[
ACR = PEN\% \times SAL_{R-1} \times a(x)
\]

52
where:

\[ \text{PEN\%} = \text{Pension received as a percent of the individual's final salary}; \]
\[ \text{SAL}_{R-1} = \text{Salary received in the final year prior to retirement}; \]
\[ a(x) = \text{Inflation-linked lifetime annuity factor for the retiree age } x \text{ (i.e., at retirement)}. \]

Equations (1) and (2) must equal each other, as the accumulated amount at retirement provides the funds required to purchase a pension at a rate related to the person's age and sex. In any individual case, however, there are two parameters: namely \( K \) (the rate of contribution) and \( \text{PEN\%} \) (the pension received in terms of final salary). Within a defined contribution fund, \( K \) is defined and the pension can be calculated based on the accumulated funds at retirement. In contrast, within a defined benefit pension fund, the pension percentage is defined (normally ignoring any tax on the pension) so that a recommended rate of \( K \) can be calculated using actuarial principles.

The above equations do not make any allowances for taxes on benefits (which vary by country, individual income, and benefit form) or expenses which may be in respect of initial expenses, regular administration or investment costs, or the costs associated with the purchase of an annuity. The important impact of expenses and the varied form in which they are paid will be considered in a subsequent study.

The provision of retirement income from savings in the preretirement years requires funds to be accumulated over many years; several long-term assumptions are therefore necessary. One approach is to use a deterministic approach and set pre-determined levels of inflation and investment return for each year. Such an approach, however, does not allow analysis of the risk facing the individual member. To provide greater reality in this model, simple stochastic models for inflation and the investment return will be used.

2.2 Inflation and Salary Assumptions

The stochastic model used for inflation allows for a one year lag as expressed in equation (3).

\[ \text{INFL}_t = k \times \text{INFL}_{t-1} + (1-k) \times (\mu + \sigma Z_t) \]  

(3)

where:
\[ Z_t = \text{Standard normal variable for year } t, \text{ i.e., } Z \sim N(0, 1); \]
\[ \text{INFL}_t = \text{Rate of inflation in year } t; \]
\[ k = \text{A number between 0 and 1;} \]
\[ \mu = \text{Mean of the normal distribution representing inflation;} \]
\[ \sigma = \text{Standard deviation of the normal distribution representing inflation.} \]

The appropriate levels for \( k, \mu, \) and \( \sigma^2 \) can be debated. After some empirical investigation into the inflation levels over the last 40 years in Australia, the following values provided a distribution of inflation values that is similar to the previous 40 years' experience:

\[ k = 0.5 \] (that is, 50 percent of last year's inflation is carried into this year);
\[ \mu = 0.07; \]
\[ \sigma = 0.07 \] (that is, the standard deviation).

The period of 40 years was chosen to cover the post-World War period. In addition, beyond 40 years there is a problem with the availability of reliable and consistent data.

Statistical tests show a significant effect for a one year lag, but no significance for a longer lagged effect. The value of \( k \) also was tested for all values between zero and one—a value of 0.5 provides a slightly better result than other values in the range of 0.25 to 0.75 and much better results than values outside this range.

Before proceeding, it is worth noting that the history of inflation does not necessarily indicate future levels. In particular, most OECD (Organization of Economic Co-operation and Development) nations have moved into a lower inflation environment. With this in mind, the results will concentrate on \( \mu = 0.04 \) and \( \sigma = 0.04 \). The effects of higher inflation rates, however, will be considered also.

As indicated above, the model requires an assumption in respect to a person's salary in each of his or her preretirement years (that is, the pattern of the person's salary from entry into the work force until retirement age). This paper's approach is to consider that the annual change in a person's salary comprises the following three components:

a) An increase related to inflation levels, which can be estimated from the inflation equation outlined above.

b) An increase as a result of general productivity improvements within the economy, which may be expressed as a percentage rate per annum.
c) A promotional increase that also will be expressed as a percentage rate per annum.

2.3 Investment Returns

The assumption of a single investment rate of return for a period of 20, 30, or 40 years to estimate the accumulated value of a person’s retirement benefit is a bold and heroic assumption and is almost certain to be wrong! To provide greater understanding of the range of possible results, each simulation assumes that each year’s rate of investment return is selected randomly from a distribution that represents the assumed experience, thereby allowing investment returns to vary on a year to year basis.

It is assumed that the investment return is achieved by a fund invested in a range of marketable assets with no promise of a guaranteed return. For the purposes of this study, it is assumed that the fund will invest in portfolios of bonds (both domestic and overseas), equities (both domestic and overseas), direct property, and short-term investments. That is, the fund will have a balanced investment strategy spread over several sectors. Naturally, the actual proportions in each sector will vary with the investment strategy adopted.

It also will be assumed that the real rate of investment return in year $t$ is independent from the rate of inflation in that year. Although this result may appear surprising, a diversified portfolio with several sectors represented is more likely to achieve this independence than a portfolio concentrated in one asset form. For instance, if inflation rises, the prices of domestic bonds will decrease and equities and property may fall in value. Short-term and overseas investments may increase in value. Carter (1991), in the development of an Australian stochastic investment model, suggests that inflation affects short-term rates positively but dividend yields and property returns negatively and that share prices best are forecast as a separate white noise process independent from inflation.

Hence, in view of the assumed diversified nature of the investment portfolio and the lack of a clear relationship between the returns on equities and inflation, a real rate of return independent from the rate of inflation is considered reasonable. It is acknowledged that this investment model is a simplified one, but it is sufficiently realistic to enable this paper to concentrate on the benefits arising from defined contribution funds and thereby to draw appropriate conclusions. Models that concentrate on interest rates (for example, Becker (1991) and Tilley (1992)) have not been used due to the assumed diversified portfolio of the fund.
It also is recognized that pension and superannuation funds may adopt a range of investment strategies. With this in mind, the results allow for the following three investment strategies, each of which is represented by a normal distribution.

a) Strategy A: \( N(\mu = 0.05 \text{ and } \sigma = 0.08) \).

b) Strategy B: \( N(\mu = 0.03 \text{ and } \sigma = 0.05) \).

c) Strategy C: \( N(\mu = 0.01 \text{ and } \sigma = 0.02) \).

It should be noted that these three investment strategies represent, in broad terms, the following three investment options:

a) Strategy A represents a managed or balanced fund with significant investments in equities and properties.

b) Strategy B represents a capital stable fund with significant fixed interest investments and some equity investments.

c) Strategy C represents a fund invested predominantly in cash and short-term stocks.

The appropriateness of the assumed figures is confirmed by Humphreys and Newman (1993) who allow for an investment mix of cash, bonds (Australian and overseas), equities (Australian and overseas), property, and currency each with its own sector statistics and show a mean return (in excess of inflation) of 5.1 percent per annum with a standard deviation of 8.2 percent for a fund with a balanced asset mix and a mean of 3.9 percent per annum real and a standard deviation of 4.8 percent for a fund with a stable asset mix. Further, the Towers Perrin Superannuation Pooled Funds Survey (1993) of Australian fund managers shows for the three years to June 30, 1993 standard deviations of 5.8 percent, 7.6 percent, and 8.9 percent per annum for the benchmarks for funds that have below average, average or above average volatility for their investment returns.

Within the model, the rate of return each year is calculated so that 1 plus the nominal rate of return in year \( t \) is the product of 1 plus the inflation rate for year \( t \) and 1 plus the real rate of return for year \( t \), for the given investment strategy. It is possible for the nominal rate of return in a particular year to be negative due to a negative real rate of return for that year.

As will be shown later, this model also permits individuals to change their investment strategies during their preretirement years, which is similar to the concept of age phasing discussed in Kingston, Piggot, and Bateman (1992). This possibility raises the question as to
who directs the investment policy: the employer, the member, or the trustees of the fund. A discussion of the advantages of each alternative is beyond the scope of this paper but is worthy of further research.

3 Results

As indicated above, the model can assume a defined contribution or a defined benefit approach. This paper initially will consider the retirement income benefits that arise for a single male in his retirement from a defined contribution of 12 percent of salary throughout his career. It is assumed that the full accumulated benefit at retirement is converted into an inflation-linked lifetime annuity.

Table 1 presents the results based on the following assumptions, except where an alternative assumption is noted.

Basic Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry age:</td>
<td>20</td>
</tr>
<tr>
<td>Exit age:</td>
<td>65</td>
</tr>
<tr>
<td>Participation:</td>
<td>full time throughout</td>
</tr>
<tr>
<td>Inflation rate–mean:</td>
<td>4 percent per annum</td>
</tr>
<tr>
<td>Inflation rate–standard deviation:</td>
<td>4 percent per annum</td>
</tr>
<tr>
<td>Investment strategy A–mean:</td>
<td>5 percent per annum real</td>
</tr>
<tr>
<td>Investment strategy A–standard deviation:</td>
<td>8 percent per annum</td>
</tr>
<tr>
<td>Investment rate after retirement:</td>
<td>1 percent per annum real</td>
</tr>
<tr>
<td>Salary growth–productivity:</td>
<td>1 percent per annum</td>
</tr>
<tr>
<td>Salary growth–promotion:</td>
<td>1 percent per annum</td>
</tr>
<tr>
<td>Mortality after retirement:</td>
<td>Australian Life Tables 1985-1987</td>
</tr>
</tbody>
</table>

The investment rate of return after retirement has been assumed to be 5 percent per annum (i.e., 1 percent in excess of the mean long-term inflation rate), as it is assumed that the institution offering the indexed lifetime annuity will adopt a more conservative investment strategy than in the preretirement period.

Table 1 indicates the spread of results that arise from 1,000 simulations undertaken for each set of assumptions by showing the mean, standard deviation, the 5th percentile, and the 95th percentile for the 1,000 results produced under each scenario. One thousand simulations is sufficient to produce a stable set of results.
The most important result shown in Table 1 is the significant spread of the level of retirement income received by individuals who have contributed the same percentage of salary for the same number of years. For instance, using the base assumptions the average retirement income arising from a contribution of 12 percent of salary for 45 years is an indexed lifetime annuity equivalent to 84.66 percent of the person’s final salary. Due to the uncertain investment returns
achieved each year, however, there exists a considerable spread of results. The level of retirement income is equally likely to be 46 percent or 143 percent of final salary, and these are not the extreme values! Figure 1 shows the distribution of these results.

![Figure 1](image)

**Figure 1**  
Retirement Income as a Percentage of Final Income

The model also allows for taxation on contributions and/or investment earnings. Table 2 shows the results assuming a 15 percent tax on contributions (which is the tax rate payable in Australia on employer contributions) and a 7.5 percent tax on investment income. This represents a typical investment income tax rate paid by funds in Australia after allowing for dividend imputation and other credits. The tax on the resulting benefits also is reduced, but this is not shown as these tax rates vary by income and benefit size.

The major message coming from the results in Tables 1 and 2 and Figure 1 is that a considerable variation occurs in the ultimate level of retirement income received by individuals, even if a level contribution rate is assumed to be paid for 40 or 45 years. In essence, a system that defines a set level of contributions cannot define the level of benefits received. With the trend toward defined contribution plans, it is critical that fund members, employers, and policy makers appreciate that the prescribed level of contributions will not provide sufficient retirement income for many retirees, even if, on average, it is satisfactory under certain circumstances. It is worth stressing that
TABLE 2
Indexed Retirement Income That can be Purchased With a 12 Percent Contribution Rate After Allowing for a 15 Percent Contributions Tax and a 7.5 Percent Investment Tax

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Retirement Income Expressed as a Percentage of Final Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Base assumptions</td>
<td>61.10</td>
</tr>
<tr>
<td>Female</td>
<td>48.94</td>
</tr>
<tr>
<td>Married male with spouse</td>
<td>44.84</td>
</tr>
</tbody>
</table>

Changes in entry or exit ages

<table>
<thead>
<tr>
<th>Retirement age</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>5th Percentile</th>
<th>95th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>41.97</td>
<td>12.87</td>
<td>24.44</td>
<td>66.44</td>
</tr>
<tr>
<td>55</td>
<td>28.74</td>
<td>8.24</td>
<td>17.61</td>
<td>44.80</td>
</tr>
<tr>
<td>Ages of 25 and 60</td>
<td>34.24</td>
<td>9.70</td>
<td>21.54</td>
<td>52.05</td>
</tr>
<tr>
<td>Age 60 (female)</td>
<td>34.27</td>
<td>10.61</td>
<td>19.76</td>
<td>54.27</td>
</tr>
<tr>
<td>Age 55 (female)</td>
<td>24.00</td>
<td>10.61</td>
<td>14.66</td>
<td>37.42</td>
</tr>
<tr>
<td>Ages of 25 and 60 (female)</td>
<td>27.96</td>
<td>8.01</td>
<td>17.36</td>
<td>42.49</td>
</tr>
</tbody>
</table>

Changes in Investment assumptions or strategy

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>5th Percentile</th>
<th>95th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>A with $\sigma = 6%$</td>
<td>61.12</td>
<td>15.04</td>
<td>40.39</td>
<td>89.17</td>
</tr>
<tr>
<td>B</td>
<td>39.95</td>
<td>7.81</td>
<td>28.73</td>
<td>54.36</td>
</tr>
<tr>
<td>C</td>
<td>27.29</td>
<td>2.73</td>
<td>23.08</td>
<td>32.19</td>
</tr>
<tr>
<td>A for 35 years, then B</td>
<td>52.05</td>
<td>10.56</td>
<td>33.34</td>
<td>78.54</td>
</tr>
<tr>
<td>A for 35 years, then C</td>
<td>44.28</td>
<td>10.61</td>
<td>29.90</td>
<td>64.27</td>
</tr>
<tr>
<td>A for 25 years, then B for 10 years, then C</td>
<td>39.01</td>
<td>7.55</td>
<td>28.58</td>
<td>52.39</td>
</tr>
</tbody>
</table>

Changes in participation rates (part time is considered 40% of full time)

<table>
<thead>
<tr>
<th>Participation Rate</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>5th Percentile</th>
<th>95th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>ft to age 30, then pt to age 40, then ft</td>
<td>51.87</td>
<td>16.29</td>
<td>30.32</td>
<td>81.08</td>
</tr>
<tr>
<td>ft to age 25, then zero to age 35, then pt to age 45, then ft</td>
<td>35.68</td>
<td>9.89</td>
<td>22.40</td>
<td>53.92</td>
</tr>
<tr>
<td>ft to age 30, then pt to age 40, then ft (female)</td>
<td>41.55</td>
<td>13.16</td>
<td>24.25</td>
<td>65.67</td>
</tr>
<tr>
<td>ft to age 25, then zero to age 35, pt to age 45, then ft (female)</td>
<td>28.58</td>
<td>8.01</td>
<td>17.83</td>
<td>43.73</td>
</tr>
</tbody>
</table>

Changes in inflation and annuity assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>5th Percentile</th>
<th>95th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflation N(4%,6%)</td>
<td>61.29</td>
<td>20.84</td>
<td>34.56</td>
<td>99.87</td>
</tr>
<tr>
<td>Inflation N(7%,7%) with annuity at 8% per annum</td>
<td>58.89</td>
<td>19.79</td>
<td>33.46</td>
<td>97.11</td>
</tr>
<tr>
<td>Annuity at inflation +0%</td>
<td>56.15</td>
<td>18.13</td>
<td>32.18</td>
<td>90.13</td>
</tr>
<tr>
<td>Annuity at inflation +1%</td>
<td>60.18</td>
<td>19.40</td>
<td>34.53</td>
<td>96.58</td>
</tr>
<tr>
<td>Annuity at inflation +2%</td>
<td>64.35</td>
<td>20.72</td>
<td>33.96</td>
<td>103.09</td>
</tr>
</tbody>
</table>

This inadequacy most likely will occur for a particular generation or cohort of retirees and not for retirees from a particular plan. For example, if the economy is depressed for a number of years (causing reduced investment returns), then all members of defined contributions plans will be affected. The effects could be particularly adverse for those approaching retirement who may find that the real value of their accumulated retirement benefits is declining. Such a result could
lead to a cohort of retirees with lower living standards for their retirement.

This result is not surprising when one recalls that within the operation of a defined contribution plan the rate credited to the member's account each year normally is linked to the fund's actual investment performance. Although funds in some countries may choose to smooth this rate, there is no doubt that the final benefit received by the individual largely is determined by the investment performance of the fund during the individual's working career. The investment risk that is borne by members of defined contribution plans translates into a risk that affects postretirement living standards. This is in contrast to a defined benefit fund where the retirement benefit is defined in terms of final (or final average) salary and the employer's contribution rate normally is adjusted to reflect changes in the investment return.

One method to reduce the variability in the level of retirement income received by the individual is the adoption of an investment strategy with less volatility, as assumed for Strategies B or C. While such an approach reduces the variability in the ultimate level of income, as shown in Table 1, a reduction in the level of retirement income also occurs. It is worth noting that, based on the model used, the 95th percentile for the low risk Strategy C represents a lower income than the 5th percentile for the higher risk Strategy A option. Similar results would be expected if other investment models were used.

A commonly suggested alternative is for individuals to reduce their level of investment risk as they approach retirement. Tables 1 and 2 show that while such a move reduces the variability in the level of retirement income received, it also reduces the expected income to be received. The expected income for the strategy involving the three investment options is below the 5th percentile for Strategy A. This result does not mean that a policy to reduce the volatility of investment return is inappropriate as individuals approach retirement. It does mean that the likely impact of such a move on the resulting income must be recognized.

The results also highlight the importance of realistic assumptions in any modeling, including variations in the rate of return. It is interesting to note that if the variations in inflation and investment returns are removed, the level of retirement income is 84.97 percent of final earnings (close to the mean). Such a single figure provides no indication of the variability in the likely results, however.

Table 1 also confirms the following results:
a) Early retirement causes a significant reduction in the level of retirement income due to the shorter accumulation period and the extended period of retirement. A retirement age of 60 causes a 32.7 percent reduction for males and a 31.4 percent reduction for females. These significant reductions in the level of retirement income need to be appreciated, particularly with recent worldwide trends toward earlier retirement.

b) Later entry into the work force, as is occurring with higher levels of youth unemployment and increasing years of education, also results in a lower level of retirement income due to the shorter period of accumulation. This reduction can be offset if the increased period of education raises the level of lifetime earnings.

c) These two trends, of later entry and earlier retirement, can have a devastating effect on the ultimate level of benefit. For instance, the expected retirement income with an entry age of 25 and a retirement age of 60 is 54 percent for males and 55 percent for females of the income received by a person who enters at age 20 and retires at age 65.

d) Changes in the investment strategy have the expected result with higher variability if the risk (as measured by the standard deviation) is increased and a reduced mean and variability if more conservative investment options are chosen. If Strategy C is chosen, the mean retirement income is reduced 59 percent while the standard deviation is reduced 88 percent.

e) If the standard deviation for Strategy A is reduced (which may occur within a prolonged low inflation environment and/or with greater smoothing of the investment returns), the expected value is almost unchanged, whereas the standard deviation and the range between the 5th and 95th percentiles are both reduced 25 percent.

f) Female life expectancy is considerably higher than males. Based on the Australian Life Tables 1985-1987, a 65 year old female is expected to live 18.56 years (or 27.1 percent longer than a male). When the retirement benefit is expressed in terms of a lifetime annuity, females receive a smaller level of income for the same level of contributions. Using the base assumptions, the expected level of income for a 65 year old female retiree is 20.0 percent below her male counterparts. (Within the Australian context, gender-based annuity rates are permitted.)

g) The previous discussion relates only to full-time workers. As expected, those who experience some periods of part-time work or who temporarily leave the work force have reduced retirement incomes. For instance, working in a part-time capacity for ten years from age 30 reduces the expected retirement income 15.5 percent for both males and females. Naturally, larger reductions in the expected retirement income occur if the person spends more time out of the work force.
h) Changes to the assumed mean of the inflation level do not cause a significant change to the results, as the investment returns and salary increases are adjusted automatically. As expected, an increase in the standard deviation of the inflation distribution leads to an increased variability in the level of retirement income.

i) Table 2 highlights the impact of a 15 percent tax on contributions and a 7.5 percent tax on investment income. The expected level of benefits is reduced 27.8 percent for both males and females under the base conditions. Even if there is a reduction in the taxation of the retirement income (as occurs in Australia, with a 15 percent tax rebate on pensions), it is likely that the introduction of taxation during the preretirement period (which has been considered in other countries) will result in a reduction in the actual level of retirement income received by the retiree.

j) The annuity rates offered at retirement to convert the accumulated benefit to a lifetime annuity can have a significant impact on the ultimate level of retirement income. If the underlying interest rate used to determine the indexed annuity rate increases from 1 percent above the inflation figure in the year preceding retirement to 2 percent above this inflation rate, the expected level of the annuity increases 6.9 percent.

This last result is important for members of defined contribution plans, as the actual level of any lifetime annuity will depend on the annuity rates available at the date of conversion. This represents a one-off conversion. The annuity rate used becomes critical in determining the actual level of retirement income received if the lump sum benefit is to be converted into an annuity stream at the date of retirement. In reality, such a system represents a random event, within certain bounds, where the level of retirement income can vary significantly due to the actual date of retirement even when all other factors are identical. Most members of defined contribution plans are not aware of this annuity rate risk that they bear whenever their retirement benefit is not expressed as a defined pension in terms of final (or final average) salary and they are required to convert their accumulated benefit into an income form. One approach to overcome this problem is to provide retirees with a fixed period (say, five years) during which they must convert their benefit into an annuity. Such an approach removes the one-off option while maintaining the requirement to convert their accumulation into an income stream.

The above results highlight the risks associated with variable investment returns and the provision of retirement benefits through a defined contribution approach. These risks cannot be removed without adopting a conservative investment strategy. Naturally, such a deci-
sion will result in lower benefits or higher long-term contribution rates. Neither of these results are optimal. An important but often forgotten question that needs to be addressed on a regular basis is: Who should bear the investment risk associated with the accumulation of contributions over the long term for the provision of retirement income? Should it be the employer, the individual, the government, or a combination of these parties?

This question has been answered in a variety of ways over time and in different countries. In some instances, the government (and hence the taxpayers) has removed the investment risk with the development of a generous social security system. This approach introduces other risks, including demographic and political risk. Elsewhere, defined pension schemes are common and the sponsoring employer has accepted the investment risk during the preretirement period. Even in these instances, certain risks remain. The trend toward defined contribution plans means that a higher proportion of the risks associated with the provision of retirement income is being accepted by the individual member. It may be claimed that with greater individual responsibility and a relative decline in the importance of the welfare state, this represents an appropriate response. It is also important, however, that individuals are aware of the consequences of the investment risk associated with defined contribution plans.

4 Summary and Conclusions

In recent years, there has been a shift from defined pension schemes to defined contribution schemes in several countries for a variety of reasons. This trend has placed a greater level of responsibility for retirement income on the individual member. Within this changed environment, individual members need to ask questions such as:

a) What is an appropriate level of contribution to provide security in retirement?
b) What are the major risks involved and who bears them?

The results in this paper, based on a simulation model using stochastic estimates for investment returns and the level of inflation, assist in preparing a response to these questions. The results in Tables 1 and 2 suggest that a total superannuation contribution rate of 9 percent to 10 percent of salary (assuming no taxation in the preretirement period) or 12 percent of salary (with taxa-
tion at the Australian levels) provides, on average, a reasonable retirement income in terms of final salary for a single male, assuming that the contributions have been paid for at least 40 years. It is important to stress that these figures ignore any social security benefits. Therefore, the contribution levels should be reduced where a social security pension also is received. The results also represent average results; it is likely that at some time in the future, a particular cohort of retirees who have saved for 40 or 45 years will receive an inadequate retirement income due to the variability of the investment returns during the preretirement period.

Even if we concentrate on the average result (which does not represent the total story), a 9 percent to 10 percent contribution rate (assuming the tax exempt scenario) is not sufficient for many individuals. Some of the circumstances where a higher contribution rate is needed include:

- Females who have longer life expectancies.
- Members with dependent spouses.
- Individuals who choose or are forced to take early retirement;
- Individuals who enter the work force later due to early periods of unemployment or increased education.
- Individuals who do not work full time throughout their career.

In many cases, an individual may be subject to a number of these factors (e.g., a female with some part-time work experience who retires at age 60) which would result in the need for a high contribution rate if a reasonable retirement income benefit is to be provided.

When one considers the small proportion of the work force who will be employed full time for 40 or 45 years and the variability in the investment returns over the long term, it is reasonable to conclude that a contribution rate equal to 9 percent of earnings will not provide an adequate level of retirement income for most retirees. This conclusion is strengthened by the fact that the above figures exclude any allowance for expenses.

Due to the enormous variety of individual circumstances, it is impossible to select a long-term contribution rate that will be satisfactory to everyone. In view of the current results, a total contribution rate for retirement income in the order of 12 percent of earnings (in a tax exempt environment) may be a reasonable long-term objective for many individuals. A higher contribution rate would be required if the fund were subject to taxation during the preretirement accumulation period.
References


David M. Knox
University of Melbourne
Centre for Actuarial Studies
Parkville, Victoria 3052
Australia
Discussion of David Knox's "A Critique of Defined Contribution Plans Using a Simulation Approach"

Michael Sze*

Professor David Knox is to be congratulated for this timely paper, which discusses a topic of major social and economic importance in many countries. Although the author's principal interest is Australia, the general trend of conversion from defined benefit plans to defined contribution plans has been the topic of many research projects in the United States and in Canada. Depending upon the emphasis of the research performed, different surveys have arrived at different conclusions. Participation in both types of plans has been relatively stable after 1984 according to Trends in Pensions 1992 (published by the U.S. Department of Labor, Pension and Welfare Benefit Administration 1992).

The stochastic approach toward analyzing the benefits provided by defined contribution plans is useful. Additional research may be done, however, to extend the methodology to include simulating inflation, salary increases, and each asset class separately. Statistics in the U.S. indicate that inflation is correlated negatively with many asset classes, and the impact of inflation of on different assets classes are different (cf. Sze, 1993, p. 43).

Another area worthy of further research is the impact of changing investment policy during the active career of the employee. A concept that has received wide acceptance in the U.S.A. is life cycle investment. The underlying principle of life cycle investment is straightforward. It promotes the discipline of matching the time horizon of investments to that of the retirement needs of the employee. Let us use the three most common asset classes—stocks, bonds, and cash—to illustrate the principle. Of these asset classes, statistics show that stocks have the highest expected return and the highest volatility. Cash is the most stable and has the lowest

* Michael Sze is a Fellow of both the Society of Actuaries and the Canadian Institute of Actuaries. He received his Ph.D. degree in mathematics from the Ohio State University and currently is a partner of Hewitt Associates. He is the chair of the Society of Actuaries Retirement Systems Research Committee, as well as a member of the Canadian Institute of Actuaries Investment Practice Committee.
expected return. The behavior of bonds lies between stocks and cash. For a young employee, the period to retirement is long. Thus, the investment portfolio should include more stocks. As the employee ages, retirement needs become more imminent and risk tolerance decreases. There should be a gradual shift toward fixed income investments. It would be instructive to examine the advantage of such an adaptable investment policy.

The paper mentions the risk of changing annuity purchase rate on conversion of the defined contribution balance to a stream of defined benefits upon retirement. Such a risk is genuine. One way that one may reduce such a conversion risk is to match the duration of assets to the duration of the expected benefit payment stream. Most insurance companies are heavily invested in fixed income assets. Thus, annuity purchase rates typically reflect the investment atmosphere of fixed income assets. If a life cycle investment policy is adopted, most of the investment in years preceding retirement should be in fixed income investments. Matching asset and benefit cash flow is not a difficult process. If such an exercise is performed and continually updated, the annuity purchase risk is reduced greatly. An area for further defined contribution research is to assess the impact of asset/benefit matching before retirement.

References


Michael Sze
Hewitt Associates
4110 Yonge Street
North York, ON M2P 2B7
Canada

Author's Reply to Discussion

Michael Sze has raised a number of issues worthy of further research: in particular, the effect of a more sophisticated inflation and/or investment model, changing investment policy during the life cycle and the possible matching of assets to pension liabilities to reduce the annuity rate risk.
I agree that each of these areas is suitable for further work. It must be realized, however, that as one becomes more particular with respect to life cycle decisions or investment models or policy, the results can be applied only to a proportion of pension fund members. Of course, this does not reduce the value of such research, but it does make the work more specific.

The purpose of this paper is to quantify, at least to some extent, the risks borne by members in defined contribution arrangements. Actuaries always have been aware of the different risk-takers in defined benefit and defined contribution plans. Many plan members and industry commentators, however, have no idea of the possible implications of belonging to defined contribution plans over the longer term. The results in this paper represent one way of illustrating these inherent risks to nonactuaries.
The Definition of Insurance: Implications for a Health Insurance Demand Model

Mark J. Browne*

Abstract

This paper uses data from the 1977-78 National Medical Care Expenditures Survey to evaluate five different measures of insurance: a family's expected out-of-pocket payment for medical care, the expected value of the indemnity (fee-for-service) benefits from an insurance policy for a family, the percentage of the expected loss that the insured pays, the policy premium, and the policy limit of coverage.

The study provides information that can help us understand whose insurance coverage will change significantly as a result of health care reform. For example, it shows that those with low income (such as minorities, families headed by females, and unmarried individuals) on average purchase low amounts of health insurance. These groups would benefit considerably if health care reform institutes universal coverage. Conversely, whites, families headed by males, married individuals, and those with high incomes on average have considerable health insurance coverage.

Key words: measures of insurance, out-of-pocket expenses, indemnity benefits

1 Introduction

Several different measures of insurance have been used by researchers to quantify a family's level of insurance coverage. This paper compares five different measures of insurance coverage using group insurance data from the 1977-78 National Medical Care Expenditure Survey. The measures that are used in this study were chosen because they have been used by other researchers to develop

* Mark J. Browne is an assistant professor of risk management and insurance at the University of Wisconsin—Madison. He received his Ph.D. from the Wharton School of the University of Pennsylvania in 1989. He currently is a member of the American Risk and Insurance Association, the American Economics Association, the Western Risk and Insurance Association, and the Risk Theory Seminar.

1 The National Medical Care Expenditures Survey was conducted during 1977 and 1978. Data were collected from approximately 14,000 randomly chosen households throughout the United States. The richness of this data set allows the construction of the five measures of insurance used in this study. Although this data set is now 16 years old, it is the most current public data set available containing all of the necessary data for this study.
equations to predict (forecast) the demand for medical insurance. Researchers have found that medical insurance demand equations (estimated using the different measures of insurance as the dependent variable) differ significantly in terms of the amount of variation explained. In addition, the statistical significance of the explanatory variables used in the models differs across models. Further, the estimated income elasticities\(^2\) of the models vary widely.

Since the data for this study were collected, the health insurance industry has undergone dramatic change. Health maintenance organizations (HMOs) have captured a significant portion of the insurance market. Similarly, preferred provider organizations (PPOs) have grown rapidly in number and size. Managed care has become an increasingly important means of controlling health care costs. Insurers have moved from merely providing health care financing to an integral involvement in health care delivery. Insurers now regularly review the appropriateness of medical care prior to agreeing to pay for it.\(^3\)

More profound changes are expected in the future as leaders in both political parties push for reform in health care financing and health care delivery. Among the measures currently being discussed is a prohibition on most types of underwriting and a requirement that employers purchase health insurance for all employees, including those who work only part time. Mandatory purchase of health insurance coupled with community rating would alter the consumption of health insurance greatly. The health insurance market today provides a myriad of different products that reflect individual and group preferences. Depending on the form of health care reform enacted, if any, individual choice in the market may be reduced significantly.

The current study provides extensive information on the types of health insurance persons demand in the absence of a government mandate that everyone has insurance. The study focuses on traditional indemnity (fee-for-service) insurance. A major advantage of indemnity insurance is that it does not restrict the insured's choice of provider. President Clinton's proposal calls for the formation of health alliances and for each health alliance to offer a choice of at

\(^2\) Income elasticity is the ratio of the proportionate change in consumption of a good relative to a proportionate change in income, with prices held constant.

\(^3\) The data used for this study preclude an analysis of policies that incorporate managed care. Managed care may increase or decrease the value of an insurance contract to an insured. Managed care programs that limit an insured's ability to collect indemnification benefits will reduce the policy's value to the insured. Programs that contribute to the insured receiving the best possible care provide enhanced value.
least one indemnity insurance policy and one HMO policy. While
HMOs and PPOs have grown considerably during the last 20 years,
indemnity insurance continues to dominate the market.

The different ways to measure health insurance discussed in this
study provide insights into what it means to have health insurance
coverage. Depending on the definition one uses to assess health insur­
ance coverage, a particular policy may be perceived to provide either
sufficient or insufficient coverage. The demand analysis provides sig­
nificant insight into the perceived value of health insurance to vari­
ous demographic groups. The findings in this study provide valuable
information for policy/decision makers and for health insurance
industry professionals.

2 The Measures of Insurance

There are two factors that make insurance particularly difficult
to quantify. First, insurance is purchased through an aleatory con­
tract, which means that the payoff from the policy is uncertain and
subject to events in the future that may or may not occur. Second, the
probability that an insurance contract actually will pay an indem­
nity benefit to an insured will depend upon the risk characteristics of
the insured as well as the provisions of the insurance contract.
Suppose an insurer sells identical policies (for the same price) to two
individuals whom it believes to be similar risks. It follows that the
insured who in fact is a greater risk will realize greater expected
benefit from the insurance. This occurs because the insurer assumes
more risk from the individual who is a higher risk than from the
individual who is a lower risk.

Before introducing the definitions of the various measures of
insurance, it is important to give an example of one model of an
insurance policy. Let $L_{ij}$ be the actual medical expenses incurred by
insured $i$ for medical service $j$ and $B_{ij}$ be the actual indemnity benefit
paid by the insurer out of the $L_{ij}$ expense. Then individual $i$’s actual
out-of-pocket expense for service $j$ is

$$OOP_{ij} = L_{ij} - B_{ij}. \quad (1)$$

The actual form of $B_{ij}$ depends on the specifics of the insurance
policy. For example, if there is a deductible of $d_{ij}$, a coinsurance per­
centage \(^4\) of $100\alpha_{ij}\% \ (0 \leq \alpha_{ij} \leq 1)$, and an amount $M_{ij}$ beyond which the
insurance company pays for the entire excess loss, then $B_{ij}$ is given as:

\(^4\) The coinsurance percentage of $100\ \alpha_{ij}\%$ is paid by the insured. Thus if $\alpha_{ij}$ is equal to
0.2, then the coinsurance percentage is 20 percent $(100 \times 0.2\%)$. 

73
\[
B_{ij} = \begin{cases} 
0 & \text{if } L_{ij} \leq d_{ij} \\
(1-\alpha_{ij})(L_{ij}-d_{ij}) & \text{if } d_{ij} < L_{ij} \leq M_{ij} \\
(1-\alpha_{ij})(M_{ij}-d_{ij}) + (L_{ij}-M_{ij}) & \text{if } L_{ij} > M_{ij}
\end{cases}
\]

For many health plans, the deductibles and coinsurance payments are applied to the insured’s total annual expenses. So, let \( N \) be the number of different types of medical services provided, and let \( L_i \) and \( B_i \) be the aggregate medical expenses and indemnity benefits, respectively. These aggregates are given by:

\[
L_i = \sum_{j=1}^{N} L_{ij}
\]

\[
B_i = \begin{cases} 
0 & \text{if } L_i \leq d_i \\
(1-\alpha_i)(L_i-d_i) & \text{if } d_i \leq L_i \leq M_i \\
(1-\alpha_i)(M_i-d_i) + (L_i-M_i) & \text{if } L_i > M_i
\end{cases}
\]

where \( d_i \), \( 100\alpha_i\% \) (\( 0 \leq \alpha_{ij} \leq 1 \)), and \( M_i \) are insured \( i \)'s annual deductible, coinsurance percentage, and the amount beyond which the insurance company pays for all excess losses.

Some plans may have an upper limit on the annual losses that will be paid, while others may have lifetime limits on the family’s medical expenses. Some plans offer deductibles that must apply to each family member, in addition to an entire family deductible. They also may have deductibles and co-payments that vary with the medical provider utilized. For example, a certain group of physicians may have negotiated with the plan to supply services under this plan at an agreed set of charges. If an insured uses one of these physicians, the deductibles and coinsurance payments are usually lower, and the upper limit \( M \) may be lower. Regardless of the type of plan, one easily can obtain accurate estimates of an individual’s expected losses. Given the myriad of different health insurance plans, however, it is not possible to give one equation for \( B_{ij} \) or \( B_i \). In addition, for all but the very simplest of plans, the mathematical form of the expected value of \( B_i \) or \( B_{ij} \) will be complicated.
**Note:** The approach used in this paper to yield expected out-of-pocket expenses and expected indemnity benefits is to apply the plan’s benefit formula to the expected losses. This is equivalent to using the approximation $E[g(X)] \approx g(E[X])$ where $g$ is a continuous function and $X$ is a real-valued random variable. This approximation is exact if and only if $g$ is a linear function of $X$.

It must be pointed out that the medical services used in this study include outpatient physician costs, inpatient surgeon fees, hospital room and board changes, fees for diagnostic tests, prescription medicine expense, and inpatient physician visits charges. Examples of medical expenses not included in the study include home health services, treatment in an extended care facility, and hospice care. The premium and claim data were collected over an 18 month period during 1977 and 1978.

Sections 2.1 through 2.5 contain a description of the five measures.

### 2.1 Expected Out-of-Pocket Payments For Medical Care ($I_1$)

The measure $I_1$ is defined as:

$$I_1 = \sum_{j=1}^{N} E[OOP_{ij}]$$

where $N$ is the number of different medical services included in the definition and $E[OOP_{ij}]$ is the expected out-of-pocket expense for the $j^{th}$ medical service for insured $i$, as defined in equation (1).

Several authors have used this measure. For example, Farley (1985) used it in her study of underinsurance in the United States, while Francis (1984) also used it in his evaluation of health insurance policies made available to federal government employees.

As $I_1$ is defined as the expected out-of-pocket expense for medical care for a family, low values of the measure suggest extensive health insurance coverage and high values indicate little coverage. Zero corresponds to full insurance coverage. For a given insurance policy, $I_1$ increases as the family’s predicted loss increases. Similarly, for a given expected loss, $I_1$ decreases as the individual’s policy provides

---

4 If the policy has a stop-loss, an increase in loss above the stop-loss will not change this measure of insurance coverage.
greater coverage of the loss.\(^5\) A policy may provide greater coverage of a loss through a variety of policy provisions such as a lower deductible, less coinsurance, higher limits, or a broader definition of covered losses.

2.2 Expected Indemnity Benefits From Insurance (\(I_2\))

This measure of insurance, \(I_2\), is used by Browne (1989) in his study of adverse selection in the individual health insurance market. Following Browne, this measure is defined as:

\[
I_2 = \sum_{j=1}^{N} E[B_{ij}]
\]

that is, \(I_2\) is the expected indemnity benefit from the insurance policy owned by insured \(i\). The size of the expected benefit depends on the expected medical expense of insured \(i\) and the provisions of the insurance policy owned by insured \(i\).

\(I_2\) is similar to the pure (or net) premium of a single insurance policy. The pure premium is the portion of the insurance premium charged on behalf of the insured to cover the anticipated cost of claims. \(I_2\) is the expected value of benefits for an insured with a particular insurance policy and a particular set of risk characteristics. It must be pointed out that \(I_2\) differs from the pure premium of the insured’s insurance policy both to the extent that the loss prediction algorithm used in this paper differs from that used by insurance company actuaries and to the extent that the data in this study differ from that used by insurers in establishing rates.

\(I_2\) will increase as expected losses increase for any given individual or family insurance policy. Likewise, the value of \(I_2\) will increase with the depth of coverage for an expected loss of a given amount. Because this measure takes into account the risk characteristics of the insured as well as the insurance policy provisions, it incorporates more information than simply the limits of coverage. It is a more appealing measure of insurance than \(I_I\), as it represents how much insurance the policy provides rather than how little.

\(^5\) For instance, suppose a family’s expected medical expense is $250. This family will have more insurance by measure \(I_I\) if the policy has a $100 deductible than if the policy has a $200 deductible. Applying the plan’s benefit formula to the family’s expected medical expense yields \(I_I = \$100\) if the policy has a $100 deductible. \(I_I = \$200\) if the policy has a $200 deductible.
2.3 Expected Out-of-Pocket Payments/Expected Losses
(I3)

The third measure of insurance, $I3$, is the ratio of the insured’s expected out-of-pocket payments for medical services to the total cost of the medical services, that is, for insured $i$,

$$I3_i = \frac{\sum_{j=1}^{N} E[OOP_{ij}]}{\sum_{j=1}^{N} E[L_{ij}]}.$$

Clearly $I3_i$ ranges in value from zero to one. Zero corresponds to full insurance and one to no insurance.

This measure of insurance differs from the coinsurance percentages that are stated in insurance policies. Stipulated policy coinsurance percentages apply to actual losses and may vary by the type of loss. For instance, the policy may stipulate one coinsurance percentage for prescription drugs and another for ambulatory surgery.

Because measure $I3$ is the fraction of the expected loss retained by the insured, it is similar to an aggregate coinsurance percentage. Here the aggregate coinsurance percentage is defined as the portion of total medical expenses that will be paid by the insured rather than by the insurer.

One factor that makes $I3$ an ambiguous measure in certain situations is that it does not increase monotonically with the size of the expected loss. An example illustrates the problem. Consider the following situation. An insured has an insurance policy with the following provisions:

- **Deductible**: $1,000
- **Coinsurance**: Insurer pays 80 percent of loss amount above $1,000 and below $5,000. The insurer pays 100 percent of loss amounts between $5,000 and $10,000.
- **Limit**: There is no insurance coverage for loss amounts in excess of $10,000.

Let $L$ be the actual loss incurred and $B$ be the actual indemnity benefit; then:
The aggregate coinsurance percentage, \( \alpha^{(agg)} \), is given by

\[
\alpha^{(agg)} = \frac{L - B}{L} = 1 - \frac{B}{L}
\]

i.e.,

\[
\alpha^{(agg)} = \begin{cases} 
0 & \text{if } L \leq 1,000 \\
0.8 \left( L - 1,000 \right) & \text{if } 1,000 \leq L \leq 5,000 \\
3,200 + \left( L - 5,000 \right) & \text{if } 5,000 \leq L \leq 10,000 \\
8,200 & \text{if } L \geq 10,000.
\end{cases}
\]

In this case, \( I3 \) is of the same form as \( \alpha^{(agg)} \), but with \( L \) replaced by \( E[L] \). The graph below shows how measure \( I3 \) varies as the expected loss increases. Note that insurance coverage ends at $10,000. Beyond $10,000, \( I3 \) increases asymptotically to 1.00.

Suppose two persons have the individual policy described above, but they have different expected losses. Individual A has an expected loss of \( L_A = $2,667 \) while individual B has an expected loss of \( L_B = $16,400 \) such that they have the same value of \( I3 \) (0.5). Does this mean that they have the same level of insurance? \( I3 \) cannot distinguish these two cases. Further, if another individual, C, also purchases the same policy but has expected losses of \( L_C = $6,000 \) such that the value of \( I3 \) is 0.3. This implies that C had less coverage than A, which is not true.

\( I3 \), however, does not suffer this problem if the policy provides unlimited coverage or if expected losses do not exceed the limits of the policy.
2.4 The Insurance Premium (I4)

The fourth measure of insurance, I4, is the premium paid for insurance coverage by or on behalf of the insured. This measure of insurance has been used widely by insurance researchers (for example Farley and Wilensky (1983) and Beenstalk et al. (1988)). An advantage of this measure is that it is relatively easy to obtain and does not require as much information to calculate as the prior three measures.

In a competitive insurance market the premium not only includes expected losses, but also includes a loading for the company’s expenses, profits, and investment income. In a market that is not competitive, insurers may earn excess profits that would be included in the premium. Because the level of company expenses, interest earnings, and any excess profits do not change the amount of insurance a policy is contractually obligated to provide, but do affect the size of the premium, this is not a perfect measure of insurance. For instance, if a company is able to exercise monopoly power in one state but not another, it may charge more for an identical policy in the state where it exercises monopoly power than for the policy in the state.
where it does not exercise monopoly power. The company is not providing any more protection in the state where it charges more than in the state where it charges less. In addition, an insurance policy premium may be an inaccurate measure of coverage if market premium rates lag changes in underlying medical expenses.

2.5 Limit of Coverage (I5)

The fifth measure of insurance, I5, is the limit of coverage specified in the insurance policy agreement. This measure of insurance is used by Phelps (1973). The limit of insurance specified in the insurance policy agreement is a characteristic of the insurance agreement and does not reflect the risk characteristics of the insured explicitly. Insurance policies often have more than one limit. Among the different limits that may be used as a measure of the coverage a health insurance policy provides are the number of hospital days covered and the maximum payment for a surgical procedure. The limit used in this paper is the lifetime maximum dollar limit of total benefits provided by the insurance policy.

I5 is defined here as the maximum amount of insurance protection that a policy will provide. The measure does not account for the differences in risk exposure inherent to different insureds. The measure also may give an imprecise indication of the amount of insurance a policy provides because it focuses exclusively on one policy provision to the exclusion of all others. Further, it is unlikely that an insured will reach the policy limit. With I5, two insurance policies with the same limit but different deductibles would not be distinguishable.

2.6 A Comparison of the Measures of Insurance

Table 1 reports the mean, standard deviation, minimum and maximum insurance coverage for the 1977-78 sample population used in this study, employing the five measures of insurance.

A procedure developed by Duan et al. (1983) was modified to predict medical expenses. Insurance policy provisions of the insureds in the sample are applied to the predicted losses to determine the value of insurance company indemnity benefits and the amount of the loss that is borne by the insured. Predicted expenses and insurance

---

6 There is little evidence that health insurers currently exercise monopoly power in the U.S. Further, health insurers' profitability is limited in many states by minimum loss ratio requirements.

7 The average medical care cost for an individual for 1994 is roughly $2500.
indemnity benefits are summed across family members to determine totals for families. *Ex ante* predicted values for the random variables *I*1, *I*2, and *I*3 then were constructed.

In the Duan et al. study, aggregate medical expenses were predicted. In the current study, losses are predicted for each person in the sample for each of six types of medical expenses: outpatient physician fees, inpatient surgeon fees, inpatient physician (nonsurgeon) fees, hospital room and board charges, fees for diagnostic tests, and prescription medicine expenses. Health insurance policies typically have different coverage provisions for different types of losses. The number of medical services for which loss amounts were predicted was limited by the availability of medical insurance data in the National Medical Care Expenditures Survey. The survey provides extensive insurance coverage information for the medical services used in the study but not for all medical services.

A simple example illustrates how the different measures of insurance coverage can conflict. Consider two families, one in good health, the other in poor health. The two families have almost identical insurance policies that cover all medical expenses after a deductible. The deductible on the policy held by the healthy family is $50. The family that is in poor health has a policy with a $100 deductible. Expected losses are $75 for the healthy family and $1,000 for the less healthy family. Suppose the insurance premium charged on behalf of each family is 120 percent of its expected loss. The expected loss is assumed to be known by the insurer. Measure *I*1 for the healthy family is $50 and $100 for the less healthy family. With *I*1 as the measure of insurance, the healthy family has more insurance than the less healthy family because *I*1 is less. A cursory

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I</em>1</td>
<td>236.94</td>
<td>210.47</td>
<td>0</td>
<td>2052.20</td>
</tr>
<tr>
<td><em>I</em>2</td>
<td>417.99</td>
<td>450.58</td>
<td>0</td>
<td>5546.10</td>
</tr>
<tr>
<td><em>I</em>3</td>
<td>0.41</td>
<td>0.27</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td><em>I</em>4</td>
<td>1012.95</td>
<td>660.90</td>
<td>21.0</td>
<td>5798.20</td>
</tr>
<tr>
<td><em>I</em>5</td>
<td>262,967.58</td>
<td>315,505.27</td>
<td>1,000.00</td>
<td>5,000,000.00</td>
</tr>
</tbody>
</table>

*I*1: Out-of-pocket payments for medical care
*I*2: Indemnity benefits from insurance
*I*3: Out-of-pocket payments/expected losses
*I*4: Insurance premium
*I*5: Limit of coverage

81
examination of the policies would lead to the same conclusion, as the healthy family’s policy is identical in all regards to that of the less healthy family except that it has a lower deductible.

Now consider how these families compare by the other measures of insurance. If measure 12 is considered as the appropriate measure of insurance, the less healthy family has more insurance than the healthy family has. The expected insurance indemnity benefit for the healthy family is $25, whereas for the less healthy family the expected benefit is $900. By measure 13, the percent of the total expected losses covered by insurance, the less healthy family has more insurance than the healthy family. The insurer will pay 90 percent of the less healthy family’s medical expenses, but only 33 percent of the healthy family’s losses. With 14 as the measure of insurance, the less healthy family has more insurance than the healthy family. Recall that premiums are assumed to be 120 percent of expected medical expenses for both families and the expected medical expense of the less healthy family is greater than that of the healthy family. Finally, with 15 as the measure of insurance, both families have an equal amount of insurance because both have a policy that provides unlimited coverage above the deductible. A summary of the example is provided in Table 2.

3 The Medical Expense Insurance Demand Model

This section of the paper specifies and estimates medical expense insurance demand equations for each of the five measures of insurance.

---

**TABLE 2**
An Illustrative Example of the Insurance Measures

<table>
<thead>
<tr>
<th>Policy Provisions</th>
<th>Healthy Family</th>
<th>Sickly Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deductible</td>
<td>$50</td>
<td>$100</td>
</tr>
<tr>
<td>Expected Losses</td>
<td>$75</td>
<td>$1,000</td>
</tr>
<tr>
<td>Limit of Coverage</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Measures of Coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:</td>
<td>$50</td>
<td>$100</td>
</tr>
<tr>
<td>12:</td>
<td>$25</td>
<td>$900</td>
</tr>
<tr>
<td>13:</td>
<td>67%</td>
<td>10%</td>
</tr>
<tr>
<td>14:</td>
<td>$90</td>
<td>$1,200</td>
</tr>
<tr>
<td>15:</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
</tbody>
</table>

11: Out-of-pocket payments for medical care
12: Indemnity benefits from insurance
13: Out-of-pocket payments/expected losses
14: Insurance premium
15: Limit of coverage
3.1 Specification of the Demand Model

For any of the measures of insurance (I) defined above, it is assumed that the medical expense insurance demand equation can be written, for family $t$, as:

$$I_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$  

Here $X_{1t}$ is a column vector of explanatory variables pertaining to the demographic characteristics of the family believed to affect the demand for insurance, and $X_{2t}$ is a vector of explanatory variables pertaining to the employment group through which the family obtains insurance coverage. Recall that the insurance demand equations are estimated with group market data. $\beta_1$ and $\beta_2$ are vectors of estimated coefficients corresponding to $X_1$ and $X_2$, respectively. The model's error term, $\epsilon_t$, is assumed to be normally distributed.$^8$

The family's demographic variables in the insurance demand equation include its perceived health status, age of the head of household, sex of the head of household, marital status, level of education of the head of household, race of the head of household, family income from all sources, family size, and region of the country. The group characteristic variables include group size$^9$ and the share of the insurance premium paid by the employer.

The family's perceived health status is hypothesized to be correlated to the amount of insurance purchased. Theoretical studies of the demand for insurance in a market with adverse selection support this hypothesis (Rothschild and Stiglitz, 1976; Wilson, 1977).

Phelps (1973) as well as others have found that age is correlated positively with the amount of medical insurance purchased. One possible explanation for this correlation is that as one ages the variance of losses increases, as does the average expected loss size.

The variable, female, has a value of one if the head of household is a female and a zero if a male. Prior studies have shown that families headed by females typically have less medical insurance than families headed by males (Farley and Wilensky, 1983).

Marital status is believed to be correlated positively with insurance purchases. Because medical expense insurance often is offered as an employee benefit extending to other members of the immediate

---

$^8$ A log transformation of the insurance measure was used in all but the I3 models. In that model the log-odds transformation was used.

$^9$ Group size is the size of the employment group through which the individual acquires insurance coverage.
family, the assumption is made that a married individual will be more likely to have medical expense insurance coverage.

Education is hypothesized to be correlated positively with insurance consumption. Grossman's (1972) model of health care consumption assumes that consumption of health care services increases one's stock of human capital. Therefore, individuals with higher levels of human capital will have a greater demand for medical insurance.

Four indicator variables are used to classify by race: black, Hispanic, white, and other. Because these indicator variables must sum to one, one must be omitted from the model. White is the omitted variable in the model. Studies of medical expense insurance consumption by race have shown that whites are the greatest consumers of medical expense insurance.

Phelps (1973) finds that income is correlated positively with insurance purchases. As insurance is not a necessity, the indigent may choose not to purchase it. As income increases, insurance becomes relatively more affordable. In addition the tax advantage to purchasing group medical expense insurance is related positively to income. This encourages a greater demand for insurance at higher income levels. A priori, the effect of income on the consumption of insurance is not clear, however. This is because those with higher incomes are more capable of self-insuring than the poor. For a risk neutral person, self-insuring may be preferable to purchasing insurance because of the expenses associated with writing insurance. Due to the extreme losses possible, self insurance usually is not opted for by persons who need to protect their assets.

Nine indicator variables are used to classify insureds by region of the country. These regions of the country are classified according to the 1970 U.S. census. Eight of the variables are included as dependent variables in the insurance demand equation. The ninth variable which represents the Pacific region is the omitted variable. These variables are included to account for regional variations in the portion of the population with medical expense insurance. They also account for the variation in the supply of medical care by region.

Five indicator variables are used to represent the area in which the insured lives. The first includes the 16 largest standard metropolitan statistical areas (SMSAs). The second includes all SMSAs (with the exception of the largest 16) that have a population greater than 500,000. The third area consists of all SMSAs with a population less than 500,000. The fourth includes areas that are not SMSAs but that are less than 60 percent rural. The fifth includes areas that are more than 60 percent rural. The first area variable,
### TABLE 3
Means and Standard Deviations For Insurance Demand Equation
Independent and Omitted Variables

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Definition</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educat</td>
<td>Years of education</td>
<td>12.26</td>
<td>3.18</td>
</tr>
<tr>
<td>Lfaminc*</td>
<td>Log of family income</td>
<td>21,342.21</td>
<td>16,554.88</td>
</tr>
<tr>
<td>Age</td>
<td>Age in years</td>
<td>39.93</td>
<td>12.26</td>
</tr>
<tr>
<td>Female</td>
<td>1 if female</td>
<td>0.17</td>
<td>0.37</td>
</tr>
<tr>
<td>Married</td>
<td>1 if married</td>
<td>0.71</td>
<td>0.45</td>
</tr>
<tr>
<td>Area2</td>
<td>1 if SMSA &gt; 500,000</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>Area3</td>
<td>1 if SMSA &lt; 500,000</td>
<td>0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>Area4</td>
<td>1 if not SMSA but &lt; 60% rural</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>Area5</td>
<td>1 if not SMSA, rural</td>
<td>0.13</td>
<td>0.33</td>
</tr>
<tr>
<td>Black</td>
<td>1 if Black</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Hisp</td>
<td>1 if Hispanic</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>Other</td>
<td>1 if not black, Hispanic, or white</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Hunk</td>
<td>1 if occupation unknown</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>Hmgr</td>
<td>1 if occupation managerial</td>
<td>0.10</td>
<td>0.31</td>
</tr>
<tr>
<td>Hsales</td>
<td>1 if occupational sales</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>Hclerk</td>
<td>1 if occupational clerk</td>
<td>0.07</td>
<td>0.22</td>
</tr>
<tr>
<td>Hcrafts</td>
<td>1 if occupational crafts</td>
<td>0.16</td>
<td>0.37</td>
</tr>
<tr>
<td>Hoper</td>
<td>1 if occupational operator</td>
<td>0.13</td>
<td>0.34</td>
</tr>
<tr>
<td>Htrans</td>
<td>1 if occupation transportation</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Hservv</td>
<td>1 if occupation service</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Hlabor</td>
<td>1 if occupation labor</td>
<td>0.05</td>
<td>0.22</td>
</tr>
<tr>
<td>Hfown</td>
<td>1 if occupation farm owner</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Hflabor</td>
<td>1 if occupation farm laborer</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>Physgood</td>
<td>1 if perceived health status good</td>
<td>0.39</td>
<td>0.49</td>
</tr>
<tr>
<td>Phsfair</td>
<td>1 if perceived health status fair</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>Phspoor</td>
<td>1 if perceived health status poor</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>Limmag</td>
<td>1 if major activity limitation</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Limmat</td>
<td>1 if amount activity limitation</td>
<td>0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>Limmin</td>
<td>1 if minor activity limitation</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>Neweng</td>
<td>1 if in corresponding 1970 census region</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>Midati</td>
<td></td>
<td>0.16</td>
<td>0.37</td>
</tr>
<tr>
<td>Encent</td>
<td></td>
<td>0.22</td>
<td>0.42</td>
</tr>
<tr>
<td>Wncent</td>
<td></td>
<td>0.08</td>
<td>0.27</td>
</tr>
<tr>
<td>Ssatl</td>
<td></td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>Seast</td>
<td></td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td>Sswift</td>
<td></td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td>Mnt</td>
<td></td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Lfamsiz</td>
<td>Log of family size</td>
<td>3.17</td>
<td>1.67</td>
</tr>
<tr>
<td>Groupszp</td>
<td>Group size</td>
<td>20,342.88</td>
<td>71,180.67</td>
</tr>
<tr>
<td>Esharec</td>
<td>Employers share of the premium</td>
<td>0.74</td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Omitted Variable</th>
<th>Definition</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Male</td>
<td>0.83</td>
<td>0.37</td>
</tr>
<tr>
<td>Area1</td>
<td>Largest 16 SMSAs</td>
<td>0.24</td>
<td>0.42</td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>0.86</td>
<td>0.35</td>
</tr>
<tr>
<td>Phsexl</td>
<td>Perceived health status excellent</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Limmon</td>
<td>No activity limitation</td>
<td>0.93</td>
<td>0.25</td>
</tr>
<tr>
<td>Pacific</td>
<td>Pacific census region</td>
<td>0.14</td>
<td>0.35</td>
</tr>
<tr>
<td>Nmarried</td>
<td>1 if not married</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>Hprof</td>
<td>Occupational professional</td>
<td>0.18</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Absolute values not log values are reported in this table. The logs of the specified variables are used in the regression.
representing the 16 largest SMSAs, is the omitted variable in the model.

Two firm-specific explanatory variables are used in the model: group size and the share of the insurance premium paid on behalf of the insured family by the employer. Group size is assumed to be a proxy for the price of insurance because group insurance can be provided more cheaply per person to members of larger groups. Thus, a positive correlation is hypothesized between group size and insurance purchases. A positive correlation also is hypothesized between the employer’s share of the premium and the consumption of medical expense insurance. For a given insurance premium, the family’s out-of-pocket cost drops as the employer’s share of the premium increases.

Table 3 reports the mean and standard deviation of the independent and omitted variables.

3.2 Estimation of the Demand Model

Table 4 reports the estimated medical insurance demand equations using each of the five different measures of insurance. Parameter estimates and tests of their significance differ markedly across equations. Of the five equations the R^2 value is highest for the I2 equation, expected insurance indemnity benefits.

Because the insurance policy premium has been used as the measure of insurance in most previous studies of insurance demand, the demand model estimated with measure I4 serves as a benchmark for comparison with other insurance demand models. The R^2 value of 0.24 falls within the range of values of other models using this measure; the R^2 value of the Farley and Wilensky model is 0.20, and that of the Phelps model is 0.39. The estimated income elasticity of 0.18 falls within the range of previously estimated income elasticities for a medical expense insurance demand model. As hypothesized, the coefficient estimates for the perceived health status variables are correlated highly with the insurance premium, as is family income. Families with a female head of household are less well insured in general than those with a male head of household. The insurance premium is related positively with family size and marital status, as hypothesized.

The age of the head of household is not significant in the demand equation. This may be attributable to the sample population being limited to individuals under the age of 65. In addition, the

---

10 Holmer (1984) provides a discussion of the variation in estimated income elasticities. Phelps estimated an elasticity of 0.18, while Hoy estimated an elasticity of 0.02.
### TABLE 4

Insurance Demand Equations

Measures of Insurance (Dependent Variables)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>I1(R²=.33) parameter estimate</th>
<th>t-stat</th>
<th>I2(R²=.39) parameter estimate</th>
<th>t-stat</th>
<th>I3(R²=.06) parameter estimate</th>
<th>t-stat</th>
<th>I4(R²=.24) parameter estimate</th>
<th>t-stat</th>
<th>I5(R²=.07) parameter estimate</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educat</td>
<td>-2.22E-11</td>
<td>-0.330</td>
<td>-0.004</td>
<td>-0.498</td>
<td>-0.004</td>
<td>-1.11</td>
<td>-0.006</td>
<td>-1.063</td>
<td>4633.133</td>
<td>1.769</td>
</tr>
<tr>
<td>LFaminc</td>
<td>0.055</td>
<td>1.961</td>
<td>0.064</td>
<td>1.780</td>
<td>0.061</td>
<td>0.401</td>
<td>0.18</td>
<td>8.283</td>
<td>16883.677</td>
<td>1.545</td>
</tr>
<tr>
<td>Age</td>
<td>0.014</td>
<td>9.638</td>
<td>0.030</td>
<td>15.540</td>
<td>-0.027</td>
<td>-3.362</td>
<td>0.003</td>
<td>2.405</td>
<td>-639.923</td>
<td>-1.106</td>
</tr>
<tr>
<td>Female</td>
<td>0.164</td>
<td>2.659</td>
<td>0.003</td>
<td>0.041</td>
<td>0.222</td>
<td>0.667</td>
<td>-0.205</td>
<td>-4.210</td>
<td>-9218.366</td>
<td>-0.400</td>
</tr>
<tr>
<td>Married</td>
<td>0.531</td>
<td>9.664</td>
<td>0.410</td>
<td>5.822</td>
<td>0.376</td>
<td>1.268</td>
<td>0.207</td>
<td>4.755</td>
<td>-12589.849</td>
<td>-0.588</td>
</tr>
<tr>
<td>Area2</td>
<td>-0.111</td>
<td>-2.090</td>
<td>-0.114</td>
<td>-1.678</td>
<td>-0.177</td>
<td>-0.617</td>
<td>-0.023</td>
<td>-0.555</td>
<td>41224.128</td>
<td>1.989</td>
</tr>
<tr>
<td>Area3</td>
<td>-0.217</td>
<td>-4.106</td>
<td>0.092</td>
<td>1.362</td>
<td>-0.801</td>
<td>-2.805</td>
<td>0.024</td>
<td>0.564</td>
<td>15951.835</td>
<td>0.773</td>
</tr>
<tr>
<td>Area4</td>
<td>-0.253</td>
<td>-4.601</td>
<td>0.031</td>
<td>0.421</td>
<td>0.012</td>
<td>0.481</td>
<td>0.012</td>
<td>-0.146</td>
<td>3269.112</td>
<td>0.152</td>
</tr>
<tr>
<td>Area5</td>
<td>-0.177</td>
<td>-2.783</td>
<td>0.015</td>
<td>0.223</td>
<td>-0.812</td>
<td>-2.732</td>
<td>0.031</td>
<td>-0.749</td>
<td>-21039.910</td>
<td>-0.847</td>
</tr>
<tr>
<td>Black</td>
<td>-0.050</td>
<td>-0.788</td>
<td>0.132</td>
<td>1.577</td>
<td>-0.133</td>
<td>-1.468</td>
<td>0.011</td>
<td>0.222</td>
<td>24025.036</td>
<td>0.962</td>
</tr>
<tr>
<td>Hisp</td>
<td>-0.226</td>
<td>-2.149</td>
<td>-0.306</td>
<td>-2.277</td>
<td>0.264</td>
<td>0.466</td>
<td>0.102</td>
<td>1.228</td>
<td>16399.984</td>
<td>0.401</td>
</tr>
<tr>
<td>Other</td>
<td>-0.171</td>
<td>-1.184</td>
<td>0.682</td>
<td>3.697</td>
<td>-1.263</td>
<td>-1.625</td>
<td>-0.105</td>
<td>-0.921</td>
<td>82126.742</td>
<td>1.463</td>
</tr>
<tr>
<td>Hunk</td>
<td>-0.126</td>
<td>-1.765</td>
<td>0.032</td>
<td>0.348</td>
<td>-0.186</td>
<td>-0.486</td>
<td>0.018</td>
<td>0.325</td>
<td>-42520.428</td>
<td>-1.534</td>
</tr>
<tr>
<td>Hmgr</td>
<td>0.087</td>
<td>1.344</td>
<td>0.183</td>
<td>2.224</td>
<td>-0.624</td>
<td>-1.798</td>
<td>-0.099</td>
<td>-1.942</td>
<td>16277.970</td>
<td>0.649</td>
</tr>
<tr>
<td>Hsales</td>
<td>0.219</td>
<td>2.558</td>
<td>0.006</td>
<td>0.055</td>
<td>-0.111</td>
<td>-0.241</td>
<td>0.046</td>
<td>0.674</td>
<td>-20834.308</td>
<td>-0.624</td>
</tr>
<tr>
<td>Hclerk</td>
<td>0.103</td>
<td>1.299</td>
<td>0.145</td>
<td>1.429</td>
<td>-0.483</td>
<td>-1.131</td>
<td>0.112</td>
<td>1.794</td>
<td>18116.812</td>
<td>0.588</td>
</tr>
<tr>
<td>Hcrafts</td>
<td>-0.022</td>
<td>-0.344</td>
<td>0.248</td>
<td>3.038</td>
<td>-0.767</td>
<td>-2.227</td>
<td>0.056</td>
<td>1.118</td>
<td>-20185.188</td>
<td>-0.812</td>
</tr>
<tr>
<td>Hopex</td>
<td>0.103</td>
<td>1.493</td>
<td>0.276</td>
<td>3.120</td>
<td>-0.683</td>
<td>-1.835</td>
<td>0.003</td>
<td>0.055</td>
<td>-56753.308</td>
<td>-2.110</td>
</tr>
<tr>
<td>Htrans</td>
<td>0.015</td>
<td>0.180</td>
<td>0.122</td>
<td>1.123</td>
<td>-0.525</td>
<td>-1.144</td>
<td>-0.004</td>
<td>-0.055</td>
<td>-12431.312</td>
<td>-0.375</td>
</tr>
<tr>
<td>Hserv</td>
<td>-0.029</td>
<td>-0.413</td>
<td>-0.373</td>
<td>4.135</td>
<td>-0.949</td>
<td>-2.497</td>
<td>0.036</td>
<td>0.651</td>
<td>-5915.009</td>
<td>-0.829</td>
</tr>
<tr>
<td>Independent Variable</td>
<td>I1 (R² = .33)</td>
<td>I2 (R² = .39)</td>
<td>I3* (R² = .06)</td>
<td>I4 (R² = .24)</td>
<td>I5 (R² = .07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>t-stat</td>
<td>Estimate</td>
<td>t-stat</td>
<td>Estimate</td>
<td>t-stat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hlabor</td>
<td>0.063</td>
<td>0.706</td>
<td>0.029</td>
<td>0.253</td>
<td>-0.324</td>
<td>-0.674</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.377</td>
<td>-1.892</td>
<td>-0.452</td>
<td>-1.769</td>
<td>-0.640</td>
<td>-0.595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hflabor</td>
<td>0.226</td>
<td>0.977</td>
<td>0.544</td>
<td>1.841</td>
<td>-0.813</td>
<td>-0.654</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phsgood</td>
<td>0.292</td>
<td>8.126</td>
<td>0.259</td>
<td>5.636</td>
<td>-0.196</td>
<td>-1.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phsfair</td>
<td>0.431</td>
<td>7.227</td>
<td>0.683</td>
<td>8.943</td>
<td>-0.762</td>
<td>-2.370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phspoor</td>
<td>0.810</td>
<td>6.359</td>
<td>1.087</td>
<td>6.666</td>
<td>-0.837</td>
<td>-1.219</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limmaj</td>
<td>0.817</td>
<td>5.589</td>
<td>0.832</td>
<td>4.443</td>
<td>1.331</td>
<td>1.688</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limmamt</td>
<td>0.420</td>
<td>4.259</td>
<td>0.938</td>
<td>7.436</td>
<td>-0.512</td>
<td>-0.963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limmin</td>
<td>0.462</td>
<td>4.274</td>
<td>0.169</td>
<td>1.226</td>
<td>0.231</td>
<td>0.397</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neweng</td>
<td>-0.145</td>
<td>-1.638</td>
<td>0.799</td>
<td>7.022</td>
<td>-1.390</td>
<td>-2.899</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midatl</td>
<td>-0.252</td>
<td>-3.762</td>
<td>0.637</td>
<td>7.431</td>
<td>-1.176</td>
<td>-3.259</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encent</td>
<td>-0.270</td>
<td>-4.451</td>
<td>0.558</td>
<td>7.198</td>
<td>-1.227</td>
<td>-3.755</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wncent</td>
<td>-0.033</td>
<td>-0.456</td>
<td>0.493</td>
<td>5.245</td>
<td>-0.910</td>
<td>-2.297</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssatl</td>
<td>-0.018</td>
<td>-0.276</td>
<td>0.267</td>
<td>3.288</td>
<td>-0.209</td>
<td>-0.613</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seast</td>
<td>-0.208</td>
<td>-2.356</td>
<td>0.021</td>
<td>0.188</td>
<td>0.103</td>
<td>0.215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swest</td>
<td>0.442</td>
<td>5.755</td>
<td>-0.320</td>
<td>-3.254</td>
<td>1.032</td>
<td>2.488</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mnt</td>
<td>0.100</td>
<td>0.988</td>
<td>-0.185</td>
<td>-1.427</td>
<td>0.940</td>
<td>1.719</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lfamsize</td>
<td>0.117</td>
<td>3.102</td>
<td>0.065</td>
<td>1.351</td>
<td>0.154</td>
<td>0.761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupszp</td>
<td>-0.014</td>
<td>-2.193</td>
<td>0.029</td>
<td>3.515</td>
<td>-0.080</td>
<td>-2.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esharec</td>
<td>-0.154</td>
<td>-2.767</td>
<td>0.145</td>
<td>2.034</td>
<td>-0.586</td>
<td>-1.955</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The log-odds transformation was used to normalize the dependent variable.
effect of age may be diluted because the ages of other individuals in
the family are not accounted for in the demand equation.

Both of the group characteristic explanatory variables are signif­
icant and exhibit the hypothesized relationship with the dependent
variable. A positive relationship between group size and insurance
consumption was hypothesized and is supported by the data. The
size of the employer portion of the insurance premium is related posi­
tively to the level of insurance purchase, as hypothesized.
The demand equations estimated using insurance measures $I_1$ and
$I_2$ exhibit the highest $R^2$ values of the five models. The measures
are related to one another by the equation:

$$I_2 = \text{Expected Medical Expenses} - I_1$$

Measure $I_1$ is a measure of the lack of insurance coverage a fam­
ily has, as it measures the family's expected out-of-pocket payments
for medical care. The interpretation of the coefficients of the demand
model estimated with $I_1$ are therefore counter to that of the other
demand models in most instances.

$I_1$ is correlated positively with the perceived health status of
the family, as is $I_2$. Those who perceive their health to be worse
demand more health insurance. The estimated coefficient for age is
statistically significant in both equations. As the age of the head of
household increases, the expected out-of-pocket payments for medical
care increase. In the $I_2$ equation the relationship is positive, indicat­
ing that the expected insurance benefits increase as the age of the
family head increases. The marital status variable is positive and
statistically significant in the $I_1$ and $I_2$ equations.

The group characteristic variables are both negative and both
significant in the $I_1$ equation. In the $I_2$ equation they are both posi­
tive and both significant. The group characteristic variables of both
equations provide strong evidence that the depth of insurance cover­
age increases with group size and with the share of the premium
paid by the employer.

The remaining two demand equations, $I_3$ and $I_5$, have low $R^2$
values for medical expense insurance demand equations. These mea­
sures theoretically are not superior to those previously discussed, and
they do not have superior explanatory power empirically.

4 Conclusion

Several different measures of insurance have been used by
researchers in previous studies of medical insurance demand to quan-
tify a family’s level of insurance coverage. This paper compares five different measures of insurance coverage using the National Medical Care Expenditure Survey data.

Insurance demand equations using the different measures of insurance as the dependent variable differ significantly from one another. The I2 equation ranks first in terms of variation explained by traditional medical insurance demand variables. Insurance demand equations utilizing measures I3 and I5 rank lowest. Explanatory variables found to be significantly related to insurance consumption in several of the equations include age, sex of the head of household, family size, income, perceived health status, and marital status.11

The information on health insurance demand provided by this research shows that the amount of health insurance persons desire varies greatly depending on demographic characteristics such as age, sex, and marital status. Income and the health of the family, proxied by predicted medical expenses, are also important determinants of the amount of insurance a family purchases. Although the leading proposals for health care reform call for community rating and universal coverage, there will likely be a sizable market for supplemental insurance after reform. This study provides valuable information on those groups that have a high demand for health insurance. In a postreform environment, the same groups likely will have the greatest demand for supplemental health insurance.

References


Farley, P.J. and Wilensky, G.R. “Household Wealth and Health Insurance as Protection Against Medical Risks.” Paper presented at the Conference on


Mark J. Browne
University of Wisconsin—Madison
School of Business
975 University Avenue
Madison, WI 53706
Discussion of Mark J. Browne’s “The Definition of Insurance: Implications for a Health Insurance Demand Model”

Charles Fuhrer*

I would like to thank Professor Mark Browne for this contribution to the actuarial literature. To the best of my knowledge, actuaries have not written on the subject of his paper.

I propose alternative measures of the amount of insurance to the author’s $I_1$, $I_2$, and $I_3$. These are:

$$I_1 = \sum_i \sum_{j=1}^N E[\text{OOP}_{ij}]$$

$$I_2 = \sum_i \sum_{j=1}^N E[B_{ij}]$$

$$I_3 = \frac{I_1}{\sum_i \sum_{j=1}^N E[L_{ij}]$$

where $i$ ranges over all insured for all insurance policies. The author selected his definitions of measures of insurance from the literature and, of course, is free to define them in any way he desires. I believe my measures would be more useful, particularly for the purpose of the paper.

My measures correspond to what would be the normal way of ordering policies. For example, suppose two policies $A$ and $B$ were identical except that they had $100$ and $200$ deductibles, respectively. For my measure, $A$ would provide more insurance than $B$. This

* Charles S. Fuhrer, FSA (1977) of the Washington National Insurance Co. in Lincolnshire, IL has been a group insurance actuary since 1973. He is co-editor of Actuarial Research Clearing House and has given numerous presentations at actuarial meetings. Mr. Fuhrer has written many papers and has been awarded the 1988 Practitioner's Prize by the AERF and the 1991 Health Section Research Papers Prize.
Defining Insurance Discussion

is the normal result because $A$ often pays more benefits than $B$ and never pays less. Under the author's definitions, if $A$ were purchased by a healthier insurance than $B$, then $A$ might have a lower value of $I2$. Generally, I would expect that a measure of the coverage level of an insurance policy would be independent of the individual that chooses to purchase it.

The author tries to answer whether certain factors influence the purchase of different amounts of insurance. Some of these factors are correlated directly with the demand for health care. Because his measures of the amount of insurance are affected by the demand for health care, correlations exist. An interesting question is: "Do individuals who will have a greater demand for health care services recognize this fact and then purchase insurance policies with greater reimbursement provisions?" I believe that the author fails to answer this question.

Charles Fuhrer
300 Tower Parkway
Lincolnshire, IL 60069-3665

Author's Reply to Discussion

The comments of Mr. Charles S. Fuhrer are greatly appreciated. His suggestion of several alternative measures of insurance is valuable. A primary aim of my paper was to raise the question of how to determine the amount of coverage an insured has. While my paper proposes several measures, future researchers may want to consider the merits of others as well.

Fuhrer's measures of an insurance policy are "independent of the individual that choose to purchase it." Such measures of insurance emphasize the contractual provisions of a policy while failing to account precisely for the amount of risk transferred by the policy. Little information on the economic value of an insurance policy to an individual insured can be gained from such measures. The measures, however, may be valuable for an insurer when examining a complete book of business.

The question raised by Mr. Fuhrer, of whether individuals who are higher risks purchase insurance policies with greater reimbursement provisions, is beyond the scope of the current study. This question has been addressed in three prior studies. Interested readers are
referred to Browne (1992), Browne and Doerpinghaus (1993), and Browne and Doerpinghaus (1994).

References


On the Equivalence of the Loss Ratio and Pure Premium Methods of Determining Property and Casualty Rating Relativities

Robert L. Brown*

Abstract

There are two distinct stages in the property and casualty ratemaking process. First, there is the portfolio average rate change. Second, there is the adjustment of classification relativities. It is well known that the loss ratio and pure premium (also called the loss cost) methods are algebraically equivalent in the stage called the portfolio average rate change. This paper reviews the proof of this equivalence. Further, it is proved algebraically that the loss ratio and pure premium methods are also equivalent in calculating classification relativities (or differentials) if certain data requirements can be met. A short numerical example of this equivalence is included.

Key words: loss cost, ratemaking, relativities

1 Introduction

In property and casualty ratemaking, there are two distinct steps in the process:

a) The portfolio average rate change.
b) A change in classification relativities.

One is able to use either a loss ratio approach or a pure premium (or loss cost) approach in these two distinct ratemaking stages. This paper first reviews the well-documented fact that the loss ratio and the pure premium approaches are algebraically equivalent when portfolio average rate changes are being calculated. The paper then

---

* Robert L. Brown, FCIA, FSA, ACAS, 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 Councilor in the City of Waterloo. Professor Brown has authored several articles and books.
proves that these methods are also equivalent when changes in classification relativities (differentials) are being calculated.¹ For these methods to be applied, the data must be in the appropriate form.

2 The Portfolio Average Rate Change

The process to be followed in developing the portfolio average rate change (also known as the statewide or provincewide rate change) is well known (see Brown, 1993, pp. 70-77) and will not be discussed directly here. There are two methods that can be used to develop rates: the loss ratio method and the pure premium method. It is relatively easy to provide mathematical formulas for these methods and to show algebraically that they are mathematically equivalent. The proof of their equivalence is well known; see, for example, Stern (1965, p. 182) and McClenahan (1990). For convenience, the proof will be repeated here. To this end, the following definitions are needed:

\[ L_{ijk} \text{ = Dollars of incurred losses for rate cell } (i, j, k); \]
\[ E_{ijk} \text{ = Units of earned exposure for rate cell } (i, j, k); \]
\[ CR_{ijk} \text{ = Current manual rate for rate cell } (i, j, k); \]
\[ i, j, k \text{ = Rating variable indicators such as } i \text{ classes, } j \text{ territories} \]
\[ \text{(There can be any number of such variables.)}; \]
\[ PLR \text{ = Permissible loss ratio } = 1 - \text{expense ratio}; \]
\[ ILR \text{ = Indicated loss ratio}; \]
\[ NAR \text{ = New average rate}; \]
\[ CAR \text{ = Current average rate}. \]

It now will be proven that the new average rate is the same for the pure premium method and the loss ratio method.

2.1 Pure Premium Method

The new average rate is determined under the pure premium method as:

\[ NAR = \frac{\sum_{ijk} L_{ijk}}{\sum_{ijk} E_{ijk}} \times \frac{1}{PLR}. \]

¹ Throughout this paper, the terms relativities and differentials are used interchangeably.
2.2 Loss Ratio Method

Under the loss ratio method, the new average rate is given by:

\[ NAR = CAR \times \frac{ILR}{PLR} \]

But the current average rate is determined as

\[ CAR = \frac{\sum_{ijk} CR_{ijk} \times E_{ijk}}{\sum_{ijk} E_{ijk}} \]

and the indicated loss ratio is:

\[ ILR = \frac{\sum_{ijk} L_{ijk}}{\sum_{ijk} CR_{ijk} \times E_{ijk}} \]

Thus, the new average rate is:

\[ NAR = \frac{\sum_{ijk} CR_{ijk} \times E_{ijk}}{\sum_{ijk} E_{ijk}} \times \frac{\sum_{ijk} L_{ijk}}{\sum_{ijk} CR_{ijk} \times E_{ijk}} \times \frac{1}{PLR} \]

\[ = \frac{\sum_{ijk} L_{ijk}}{\sum_{ijk} E_{ijk}} \times \frac{1}{PLR} \]

which is the same as the new average rate derived by the pure premium method.

3 Change in Classification Relativities

Again, there are two methods that can be used to change classification relativities: the pure premium (or loss cost) method and the loss ratio method. Some confusion exists, however, about which method is better and why. Also, the classical ratemaking papers found in the Casualty Actuarial Society’s associateship syllabus may not make clear what data must be used to guarantee a correct analysis.
For example, Stern (1965, p. 170) states in the section on classification relativities:

The pure premium indices above measure the relationship of the loss cost per car for each class to the base class. Consequently, they also indicate how the rate for each class should relate to the rate for the base class, if it is accepted that the expense portion of the rate is obtained by a uniform expense loading ... However, pure premiums obtained from a consolidation of widely divergent bodies of experience must be used with great caution since they may contain distortions. The above model may contain in Class 11 a proportionally larger share of experience coming from low loss cost territories than is contained in the experience for Class 12. Consequently, a part of the indicated rate differential is purely due to distribution; this distortion due to distribution would have to be corrected for, prior to accepting pure premium indices as true indications of classification relativities.

Stern (1965) continues:

There are, however, many advantages in favor of using collected loss ratios. These loss ratios can be obtained with relative ease directly from the experience; unlike pure premiums, they are less likely to be distorted by the influence of divergent distributions, since the premiums reflect the different rate and loss levels of the component territories; and finally, loss ratios based on the actual experience have an air of reality, reflecting the over-all underwriting record for each class.

Finger (1990, Chapter 5, p. 259) states:

When earned premium is used, the method is usually a "loss ratio" method; when earned exposures are used, the method is usually a "pure premium" approach. The loss ratio method can produce equivalent results if "earned premiums at current rates" are calculated.

Finger (1990) adds:

There are advantages and disadvantages of using the loss ratio and pure premium methods. The loss ratio method may be applicable when there is less detailed data available or when there are many different sets of relativities; earned premiums will reflect the various charges made for different classes, territories, and coverages. If earned premiums correspond to historical rate levels, however, it may be difficult to make adjustments for intervening changes in rate relativities. The pure premium approach is usually more accurate, because it requires more information. It also has the advantage of producing frequency and severity relativities, as well as pure premium relativities; the loss ratio method only produces loss ratio and severity relativi-
ties. Severity relativities, however, will not be meaningful if the underlying coverage is not consistent (e.g., there are differing deductibles or insured limits).

Finger then provides an arithmetic illustration of an actual calculation of some classification relativities using both the loss ratio and pure premium techniques. In his solved example using the pure premium method, Finger does not use just earned exposures for the denominator of each respective loss cost. Rather, he calculates and uses what he calls *base exposures*. He explains base exposures (p. 266):

It should be noted that "base exposures" are used in this exhibit in place of earned exposures. "Base exposures" are calculated using the current rate relativities for all relevant rating variables.

Finger argues that the reason for using base exposures instead of actual exposures is to correct for varying exposure levels in the non-reviewed relativities. For example, Territory A and Territory B may differ in the distribution of insureds by class.

Finger corrects for the distortion caused by the heterogeneity of exposure distributions across the variables not now under review, as previously alluded to in Stern's paper (e.g., varying exposure levels by class in the different territories) and for which Stern suggests corrections must be made. This is illustrated in the example in section 5 below.

Finger provides a one line arithmetic illustration of how the base exposure adjustment is made. It is difficult to conclude, however, that an average reader could reproduce the solution with only the information available.

Some questions remain: Which is superior, the loss ratio method or the pure premium method? What does Finger mean when he says that "The loss ratio method can produce equivalent results if 'earned premiums at current rates' are calculated?" Unfortunately, Finger does not elaborate further on this comment.

To deal with these questions, an algebraic description of this aspect of the ratemaking process must be developed. Without loss of generality, consideration is limited to cases where there are only two classification parameters. Define two vectors of differentials:

\[ x_i \text{ for } i = 1, 2, \ldots, n \text{ (e.g., class)} \]

\[ y_j \text{ for } j = 1, 2, \ldots, m \text{ (e.g., territory)} \]
Robert L. Brown

Assume there is a base cell, $B$, for any variable, such that for that cell:

$$x_B = y_B = 1.000$$

The current rate for the base cell will be denoted $CR_B$. Otherwise, the notation used is as defined previously.

Consider a rate manual produced by the base rate $CR_B$ and the two vectors of relativities $x_i$ and $y_j$. This produces a matrix of $m \times n$ rates. Consider (without loss of generality) that the new differential for class $k$, $x_k$ is to be calculated. One can think of class (or territory) $k$ as occupying the $k^{th}$ row of our rate manual matrix.

The calculations that follow assume that the various rate relativities are calculated independently (as opposed to interactively, as in Brown’s (1988) minimum bias approach) and applied multiplicatively. While the latter assumption is not essential in practice (i.e., additive differentials are possible), multiplicative differentials are the norm. The algebraic proofs that follow assume a multiplicative relationship. The proofs also assume that all expenses are expressed as a percentage of the gross premium (i.e., there are no flat-loaded expenses). This means that the loss relativities and rate relativities are the same.

The papers by Stern and Finger indicate that the calculation of multiplicative rating relativities can be expressed algebraically as follows:

### 3.1 Pure Premium Method

The loss cost for variable $k$, $LC_k$, adjusted for heterogeneity under the pure premium method is:

$$LC_k = \frac{\sum L_{kj}}{\sum E_{kj} y_j}.$$  

The loss cost for the base cell, $B$, adjusted for heterogeneity under the pure premium method is denoted by $LC_B$ where:

$$LC_B = \frac{\sum L_{Bj}}{\sum E_{Bj} y_j}.$$  

Thus, the new differential is:
\[ x_k \text{ new} = \frac{LC_k}{LC_B} = \frac{\sum L_{kj}}{\sum E_{kj} y_j} \times \frac{\sum E_{Bj} y_j}{\sum L_{Bj}}. \]

### 3.2 Loss Ratio Method

The loss ratio for variable \( k \), \( LR_k \), is determined as:

\[ LR_k = \frac{\sum L_{kj}}{\sum E_{kj} CR_{kj}}. \]

The loss ratio for the base cell, \( B \) is:

\[ LR_B = \frac{\sum L_{Bj}}{\sum E_{Bj} CR_{Bj}}. \]

Thus, the adjustment factor for cell \( k \), \( AF_k \), is:

\[ AF_k = \frac{\sum L_{kj}}{\sum E_{kj} CR_{kj}} \times \frac{\sum E_{Bj} CR_{Bj}}{\sum L_{Bj}}, \]

and the new differential is determined as:

\[ x_k AF_k = x_k \frac{\sum L_{kj}}{\sum E_{kj} CR_{kj}} \times \frac{\sum E_{Bj} CR_{Bj}}{\sum L_{Bj}}. \]

But \( \sum E_{kj} CR_{kj} = \sum E_{kj} CR_B x_k y_j = CR_B x_k \sum E_{kj} y_j \) (existing \( x_i, y_j \)). Thus,

\[ x_k \text{ new} = x_k \frac{\sum L_{kj}}{CR_B x_k \sum E_{kj} y_j} \times \frac{CR_B x_B \sum E_{Bj} y_j}{\sum L_{Bj}} \]

and \( x_B = 1 \). Therefore
Thus it can be seen that when the correct data are used, the pure premium method and the loss ratio method are algebraically equivalent.

4 Comments

Now that we have proved algebraically the equivalency of the loss ratio and the pure premium methods in the entire ratemaking process (i.e., both the overall rate change and also the change in relativities) if the appropriate data are used, a number of issues surrounding the calculation of risk classification relativities disappear or are resolved.

First, if the data requirements can be satisfied, then the loss ratio method and the pure premium method provide equivalent results. Therefore, there should be no need to discuss the advantages of one method over the other. They are equivalent given the appropriate data are available and used. To the extent that one cannot attain the data requirements, then one can see clearly what inadequacies will result because of the particular data one often is forced to use.

For example, if in the loss ratio approach one uses collected earned premiums (or collected loss ratios, as Stern suggests, because they are readily available), this will result in an error to the extent that the collected earned premiums are not equal to earned premiums at the current rate level. If there have been some sizable changes in relativities in recent rate changes, then this will be a problem. If the relativities have not changed drastically over the last few rate changes, however, then there may not be much of a difference between collected earned premiums and earned premiums at current rate levels. (Note that overall rate changes are not of any consequence at this stage; only the changes in classification relativities matter.)

Also, this algebraic illustration shows exactly what is meant by Finger's base exposures. These are effective exposure units that are adjusted because of the heterogeneity of exposure distributions across the different rating parameters. The following illustration makes this clear.
5 Illustration

Given the following information, and assuming the revised rates take effect July 1, 1993 for one year on one year policies, determine new rates for each of Class 1 and Class 2 and for each of Territory 1 and Territory 2. (Class differentials will not change.) Use both the loss ratio and pure premium methods. The permissible loss ratio is 0.600, and all data are fully credible.

<table>
<thead>
<tr>
<th>Present Rates:</th>
<th>Territory 1</th>
<th>Territory 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1 (Relativity)</td>
<td>100 (1.00)</td>
<td>200 (2.00)</td>
</tr>
<tr>
<td>Class 2 (Relativity)</td>
<td>300 (3.00)</td>
<td>600 (6.00)</td>
</tr>
<tr>
<td>Collected Earned Premium</td>
<td>1,000,000</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Policy Year 1991 Incurred Losses</td>
<td>360,000</td>
<td>240,000</td>
</tr>
<tr>
<td>Expected Effective Period Incurred Losses (Trended and Developed)</td>
<td>612,000</td>
<td>408,000</td>
</tr>
<tr>
<td>Earned Exposure Units:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>5,000</td>
<td>2,000</td>
</tr>
<tr>
<td>Class 2</td>
<td>1,000</td>
<td>500</td>
</tr>
</tbody>
</table>

The solution is given below for the loss ratio and pure premium methods. For each of the two methods, the rate change involves three stages: overall average rate change, change in relativities, and balance back.

5.1 Loss Ratio Method

5.1.1 Overall Average Rate Change

For the loss ratio method, the actuary must calculate the earned premium at current rates. The accounting entry for collected earned premium is not the correct denominator, because it could contain earned premiums based on the rates in out-of-date rate manuals.

The earned premium at current rates is calculated as

\[ \sum_{ij} CR_{ij} \times E_{ij} \]

\[ = (100)(5,000) + (300)(1,000) + (200)(2,000) + (600)(500) \]

\[ = 1,500,000. \]
This produces an expected effective period loss ratio at current rates of:
\[
\frac{1,020,000}{1,500,000} = 0.680,
\]
which, with a permissible loss ratio of 0.600, leads to an indicated rate change of
\[
\frac{0.680}{0.600} - 1 = + 13.3\text{ percent}.
\]

5.1.2 Change in Relativities

The given data allow for a territorial relativity change analysis but not a class relativity change analysis because loss data by class are not given. We are told that class relativities will remain the same and are asked to determine the indicated new relativities for Territories 1 and 2.

For Territory 1 the earned premium at current rates equals:
\[
(100)(5000) + (300)(1000) = 800,000
\]

For Territory 2 the earned premium at current rates equals:
\[
(200)(2000) + (600)(500) = 700,000
\]

<table>
<thead>
<tr>
<th>Territory</th>
<th>Existing Differential</th>
<th>Loss Ratio at Current Rates</th>
<th>Indicated Differential</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>360,000/800,000 = 0.4500</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>2.00</td>
<td>240,000/700,000 = 0.3429</td>
<td>0.3429/0.4500 (2.00) = 1.5238</td>
</tr>
</tbody>
</table>

Note that as presented, the Territory 1 relativity has been left at 1.00, whereas the Territory 2 relativity has been reduced from 2.00 to 1.5238. This suggests that the actuary could define the new rates as follows:
If this were done, however, the resulting rate increase would be less than the required +13.3 percent due to the off-balance created by the method used to change relativities. This is adjusted in the balance-back step.

5.1.3 Balance Back

The existing average differential is equal to:

\[
\frac{(5,000)(1) + (1,000)(3) + (2,000)(2) + (500)(6)}{8,500} = 1.7647.
\]

The proposed average differential is equal to:

\[
\frac{(5,000)(1) + (1,000)(3) + (2,000)(1.5238) + (500)(4.5714)}{8,500} = 1.5686.
\]

The balance-back factor is calculated as:

\[
\frac{\text{Existing average differential}}{\text{Proposed average differential}} = \frac{1.7647}{1.5686} = 1.1250,
\]

leading to the following proposed rates:

<table>
<thead>
<tr>
<th></th>
<th>Territory 1</th>
<th>Territory 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>113.33</td>
<td>172.70</td>
</tr>
<tr>
<td>Class 2</td>
<td>340.00</td>
<td>518.09</td>
</tr>
</tbody>
</table>

These proposed rates will result in a 13.3 percent increase in premium income, as required.

5.2 Pure Premium

5.2.1 Overall Average Rate Change

We know that the expected effective period incurred losses (developed and trended) equal 1,020,000, from which we find the indicated loss cost:
and the average rate:

\[
\frac{120}{PLR} = \frac{120}{0.600} = 200.00.
\]

Note that this is the indicated average gross rate. It is not the indicated rate for any particular territory or class that will be determined when we know the new average relativity for the expected book of business.

5.2.2 Change in Relativities

To set the new territorial relativities, the actuary normally calculates the average loss costs for Territory 1 and Territory 2 and compares them as follows:

<table>
<thead>
<tr>
<th>Territory</th>
<th>Existing Relativity</th>
<th>Loss Cost per Unit</th>
<th>Indicated Relativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>( \frac{360,000}{6,000} = 60 )</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>2.00</td>
<td>( \frac{240,000}{2,500} = 96 )</td>
<td>1.60</td>
</tr>
</tbody>
</table>

This is not the same answer as we got from the loss ratio method. Remember that the pure premium method will be correct only if the heterogeneity of distributions of exposure units is accounted for. Recall the following earned exposure unit data:

<table>
<thead>
<tr>
<th>Territory 1</th>
<th>Territory 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>5,000</td>
</tr>
<tr>
<td>Class 2</td>
<td>1,000</td>
</tr>
</tbody>
</table>

In Territory 1, 5/6 of drivers are Class 1 and 1/6 are Class 2. In Territory 2, 4/5 of drivers are Class 1 and 1/5 are Class 2. To arrive at the correct answer, this heterogeneity of cross-variable distributions must be reflected. One way to accomplish this is to use exposure units that are weighted by their cross-parameter relativities. That is, Class 1 will count as an exposure unit with weight 1.00, but Class
2 will count as an exposure unit with weight 3.00, because of its class relativity of 3.00. This leads to the following results:

<table>
<thead>
<tr>
<th>Territory</th>
<th>Existing Relativity</th>
<th>Weighted Units of Exposure</th>
<th>Loss Cost per Weighted Unit of Exposure</th>
<th>Indicated Relativity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>8000</td>
<td>360,000</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>3500</td>
<td>240,000</td>
<td>1.5238</td>
</tr>
</tbody>
</table>

5.2.3 Balance Back

Finally, the actuary determines the rate for Territory 1 and Class 1 that will produce all of the correct manual rates by balancing back for the average indicated relativity. That is:

\[
\text{Base rate} = \frac{\text{Average rate}}{\text{Average relativity}}
\]

where the average rate is 200 and the average relativity is

\[
\frac{(5,000)(1) + (1,000)(3) + (2,000)(1.5238) + (500)(4.5714)}{8,500} = 1.5686.
\]

This leads to

\[
\text{Base rate} = \frac{200}{1.5686} = 127.50.
\]

The resulting manual rates are the same as with the loss ratio method, as expected. This gives us indicated rates where all calculations are based on existing relativities in the current rate manual and should be treated as a first iteration indicated relativity. These indicated relativities will be used in a second iteration (for example, to recalculate the premium at current rate levels in the loss ratio method) to arrive at a second iteration indication. This process soon converges to the final relativities.

References


Robert L. Brown

Department of Statistics & Actuarial Science
University of Waterloo
Waterloo ON N2L 3G1
Canada
Funding Methods and Pension Plan Amendments

Keith P. Sharp*

Abstract**

This paper considers the treatment of plan amendments under the individual entry age normal and projected unit credit methods. Alternative treatments are considered, and comments are made about their acceptability.

Key words: nonretroactive amendment, normal cost, entry age normal, projected unit credit

1 Introduction

It is common for a pension plan to be amended to improve benefits in respect of service after the date of amendment. This will be referred to as a nonretroactive amendment. The application of the entry age normal and projected unit credit cost methods to this situation requires that a decision be made about the way to handle such an amendment. This paper considers these two cost methods and their application to such an amendment. A retroactive improvement can be treated in a more straightforward manner and is not considered in this paper.

The discussion of the entry age normal method is relevant to funding calculations under the Pensions Benefits Acts in Canada and under the Employee Retirement Income Security Act of 1974 (ERISA) and Internal Revenue Code (IRC) Regulation Section 1.412 in the United States. The discussion of the projected unit credit method is relevant to funding calculations and pension expense calculations.

Before developing the main results of this paper, it is important to introduce the notation used in the sequel. As there is no internationally accepted standard pension notation, we will follow, to a large extent, the notation used by Anderson (1992).

* Keith Sharp is an associate professor in the department of statistics and actuarial science at the University of Waterloo. He is a fellow of the Society of Actuaries and has a Ph.D. in finance.

** The author thanks anonymous referees for helpful comments and the Natural Sciences and Engineering Research Council of Canada for financial support.
\[ j = \text{Label of an individual member of the plan;} \]
\[ NC_i(t) = \text{Normal cost for individual } j \text{ at time } t, \text{ paid at the beginning of each year and expressed in dollars;} \]
\[ v_j = \text{Age on the first valuation date coinciding with or next following the date of participation assuming current participation requirements always had been in effect;} \]
\[ w_j = \text{Age from which credited pensionable service is calculated, i.e., the entry age for individual } j \text{ that determines the start of the period to which the benefit formula applies. In some cases the individual may join the plan after age } w_j \text{ and be given retroactive pensionable service;} \]
\[ x_j(t) = \text{Age at time } t \text{ of individual } j; \]
\[ y_j = \text{Retirement age of individual } j; \]
\[ B_i(y_j) = \text{Projected annual pension benefit of individual } j \text{ from retirement at age } y_j; \]
\[ S^j = \text{Projected measure of final pay for individual } j; \text{ and} \]
\[ s_x = \text{Salary scale for individual } j \text{ at age } x_j. \]

2 Plan Amendments Under Individual Entry Age Normal

2.1 Individual Entry Age Normal

The individual entry age normal pension cost method is used in both the United States and Canada. There are two common forms of the method (Anderson, 1992, pp. 13-19; Trowbridge and Farr, 1976, pp. 47-54; and Berin, 1989, p. 14). Under one form, the normal cost is expressed as a level dollar annual amount. This method is alternatively known as the projected benefit cost method (with supplemental liability, constant amount) (Winklevoss, 1977), the entry age actuarial cost method, and the level dollar cost method (entry age, with supplemental liability) (McGill and Grubbs, 1989, p. 27). Under another form, the normal cost is expressed as a level percentage of salary. The latter method also is known as the projected benefit cost method (with supplemental liability, constant percentage) (Winklevoss, 1977), the entry age actuarial cost method, and the level percentage cost method (entry age, with supplemental liability), (McGill and Grubbs, 1989, p. 327).

Under the individual entry age normal method, the normal cost is found by taking an equation of value. The equation usually is taken on the first valuation date coinciding with or next following a member's participation date, assuming current participation requirements always had been in effect. This age could be that at a date before
plan inception. The normal cost under the level dollar method, equation (1), is given from this equation of value by dividing both sides of the equation by the service-based annuity (Anderson, 1992, p. 13):

\[
NCi(t) = B\dot{i}(y_j) \frac{D_{y_j}}{\bar{a}_{y_j}} \times \frac{1}{\bar{a}_{y_j-y_j}}
\]

Under the level percentage of salary method (Anderson, 1992, p. 18), the annuity in the denominator of equation (1) takes the salary scale into account. The dollar normal cost is found by multiplying by the ratio of the salary scale factors:

\[
NCi(t) = B\dot{i}(y_j) \frac{D_{y_j}}{N_{y_j} - N_{y_j}}.
\] (1)

The focus of this paper is the choice of a cost method variant that is acceptable and makes sense to a client on a plan amendment; this amounts to a discussion about the method of calculating \(B\dot{i}(y_j)\). For simplicity it is assumed that all retirements occur at age \(y_j\) and that the only benefit is a retirement annuity.

### 2.2 Plan Amendment

We focus attention on \(B\dot{i}(y_j)\). For the purpose of illustration, we will assume that the benefit is a fraction \(r_0\) (e.g., \(r_0 = 0.01\) or \(0.02\)) of a projected measure \(S_f\) of final pay for each year of credited pensionable service. The measure \(S_f\) will depend on the plan document definition of the pension benefit; \(S_f\) may be, for example, the average of the earnings in the final three years of service. Thus, prior (subscript \(p\)) to any possible plan amendments, we have:

\[
B^i_p(y_j) = r_0(y_j - w_j)S^p_f.
\] (3)

From equations (1) and (2) we can see that two persons with the same entry age \(w\) and the same retirement age \(y\) will have the same normal cost as a fraction of the measure of final salary.

Now consider a situation where at a certain date, the benefit fraction \(r_0\) is changed nonretroactively from \(r_0\) to \(r_1\). Usually \(r_1\) will exceed \(r_0\), although the funding methods discussed here apply math-
ematically, if not in the view of regulators, also to the case where \( r_0 \) exceeds \( r_1 \). One method is to spread the funding of the increase over the period from amendment to retirement, with no change in the amendment date actuarial liability; this is the individual level premium method as described by Anderson (1992, p. 25). Two of the possible methods of handling this situation under entry age normal are described below.

2.3 Variant 1: EAN Total Service Spread

For an individual \( j \) with pensionable service credited from age \( w_j \) and age at plan amendment \( x_j \), one initially might assume that the projected benefit should be given by:

\[
B_{A1}^j(y) = [r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A))] s^j
\]

where the subscript \( A \) indicates that the situation after the plan amendment is being considered and \( t_A \) is the date of the amendment. This indicates that the normal cost for individual \( j \), by equation (1) and (2), would increase under this EAN—total service spread method in the ratio

\[
\frac{EANNC_{A1}^j}{EANNC_p^j} = \frac{[r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A))] \cdot r_0(y - w_j)}{r_0(y - w_j)}.
\]

This ratio depends on the values of \( x_j(t_A) \) and \( w_j \). For example, for two members \( i \) and \( k \) with the same pensionable service commencement dates \( w_i = w_k \) but differing ages at amendment \( x_i(t_A) \neq x_k(t_A) \), the normal cost as a fraction of salary no longer will be the same as the fraction of the measure of final salary. Also, the increase in the normal cost is not the same ratio \( r_1/r_0 \) as the increase in the benefit accumulation rate.

It is instructive also to consider the effect on the actuarial liability \( AL \). At age \( x_j(t) \) prior to the plan amendment, the actuarial liability is the difference between the present values of future benefits and future normal costs:

\[
EANAL_p^j(x_j(t)) = PVFB_p^j - EANPVFNC_p^j(x_j(t)).
\]

Immediately after the plan amendment at age \( x_j(t_A) \) we have (noting that the future benefits should be those actually projected to be paid for both the constant dollar and constant percentage methods):

\[
EANAL_p^j(x_j(t)) = PVFB_p^j - EANPVFNC_p^j(x_j(t)).
\]
Thus, the plan actuarial liability at the date of the amendment increases because of the amendment, although the benefit rate change is not retroactive. The proportionate increase in the actuarial liability equals the proportionate increase in the projected benefit. This aspect may be difficult to explain to a client who is not an actuary. The increase in accrued liability results because the normal cost increases only by the same proportionate amount as the increase in projected benefit. If \( r_1 < r_0 \) then the accrued liability is reduced, which may be unacceptable to regulators.

### 2.4 Variant 2: EAN Retroactive NC Mimic

An alternative method of handling normal costs under a plan amendment is described in this section. It is used by some pension consultants and gives results that are more acceptable than those described in the previous section.

Under variant 2 (EAN retroactive NC mimic), the hypothetical projected benefit is used in calculating the normal cost under this version of the entry age normal cost method. It is that projected benefit that would be applicable if the amended benefit rate were applied to all service:

\[
B_{A2}^j = r_1 (y_j - w_j) S_j^j.
\]

Under this variant, the normal cost at any post-amendment time \( t \) for individual \( j \) increases under both the level dollar and level percentage methods in the ratio of the benefit rates.
\[
\frac{NC_A^j(t)}{NC_p^j(t)} = \frac{r_1 (y_j - w_j)}{r_0 (y_j - w_j)} = \frac{r_1}{r_0}.
\] (9)

Under the individual entry age normal method the normal cost is not interpreted as being the cost of the benefit accrual for the year. Nonetheless, a proportional increase in normal cost equal to the proportional increase in benefit rates is likely to be intuitively appealing to the client.

Let us now consider the actuarial liability under variant 2. Immediately after the plan amendment, variant 2 is given for both the level dollar and level percentage methods by:

\[
EANAL_A^j(x_j(t_A)) = PVFB^j_A(x_j(t_A)) - EANPVFNC^j_A(x_j(t_A)) = \frac{r_1 - r_0}{r_0} x PVFB^j_p(x_j(t_A))
\] (10)

Because the actual future benefits are the same for variants 1 and 2, \(PVFB^j_A(x_j(t_A)) = PVFB^j_A(x_j(t_A))\). Then we note that \(PVFB^j_A(x_j(t_A))\) is related to \(PVFB^j_p(x_j(t_A))\) by the proportionate increase in the projected benefit. Also, the future normal cost increases in the ratio \(r_1/r_0\). Hence:

\[
EANAL_A^j(x_j(t_A)) = \left[\frac{r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A))}{r_0(y_j - w_j)}\right]PVFB^j_p(x_j(t_A))
\]

\[
-\frac{r_1}{r_0} x EANPVFNC^j_p(x_j(t_A))
\]

\[
= EANAL_p^j(x_j(t_A)) + \frac{(r_1 - r_0)}{r_0} x
\]

\[
\left[\left(\frac{y_j - x_j(t_A)}{y_j - w_j}\right)PVFB^j_p(x_j(t_A)) - EANPVFNC^j_p(x_j(t_A))\right]
\]

\[
= EANAL_p^j(x_j(t_A)) + \frac{(r_1 - r_0)}{r_0} x PVFB^j_p(x_j(t_A)) x
\]

\[
\left[\frac{y_j - x_j(t_A)}{y_j - w_j} - \frac{sD_{x_j(t_A)}}{sD_{w_j}} x \frac{s\delta_{x_j(t_A)}: y_j - x_j(t_A)}{s\delta_{w_j}: y_j - w_j}\right].
\] (12)
Appendix A shows that if \( r_1 > r_0 \) and if \( sD_z \) is a decreasing function of \( z \), then:

\[
EAN_{AL}^j(x_j(t_A)) \geq EAN_{AL}^j(x_j(t_A)) .
\]  
(13)

For \( r_1 < r_0 \), the actuarial liability is reduced by the amendment.

The last term of equation (11) is likely to be small; the actuarial liability is changed little by the nonretroactive amendment. This is likely to make sense to a client.

In the United States, IRC Regulation Section 1.412(c)(3)-1(c)(2) requires “If each actuarial assumption is exactly realized under a reasonable funding method, no experience gains or losses are produced.” This condition is satisfied by variant 2, as indicated in Appendix B.

2.5 Variant 3: EAN/ILP

A third method of handling the plan amendment under entry age normal is to use the individual level premium (ILP) method. This usually is regarded as a cost method in its own right; here it will be regarded more as a variant of entry age normal. The terminology EAN/ILP will be used.

Under variant 3, the nonretroactive benefit increase at \( t_A \) is funded over the period from \( x_j(t_A) \) to \( y_j \). Hence the normal cost after the amendment is given by:

\[
EAN_{NC}^j(x_j(t))
\]

\[
= EAN_{NC}^j(t) + S_j(r_1 - r_0)(y_j - x_j(t_A)) \alpha_{y_j}^{12}
\times \frac{D_{y_j}}{D_{x_j(t_A)}} \times \frac{sD_{x_j(t_A)}}{sN_{x_j(t_A)} - sN_{y_j}} \times \frac{s_{x_j(t)}}{s_{x_j(t_A)}}
\]  
(14)

and the actuarial liability at an age \( x_j(t) \), \( t \geq t_A \) is

\[
EAN_{NC}^j(x_j(t))
\]

\[
= PVFB^j_p(x_j(t)) + S_j^j (r_1 - r_0)(y_j - x_j(t_A)) \alpha_{y_j}^{12} \frac{D_{y_j}}{D_{x_j(t)}}
\]
Immediately following the amendment, the actuarial liability is found by substituting $t = t_A$ in equation (15):

$$EAN_{AL}^j(x_j(t_A)) = PVFBC^j(x_j(t_A)) - PVFN_{C}^j(x_j(t_A)),$$

Thus, as is arranged by construction of variant 3, the actuarial liability at the time of the amendment is unchanged by the amendment. Considering equations (9) and (13), it is evident that the normal cost under variant 3 must increase at the amendment by more than the ratio by which it increases for variant 2 for $r_1 > r_0$:

$$\frac{EAN_{NC}^j(x_j(t_A))}{EAN_{NC}^j(x_j(t_A))} \geq \frac{r_1}{r_0}.$$  

This behavior compares with the $r_1/r_0$ proportionate increase in normal cost under equation (9) (variant 2, the EAN retroactive NC mimic). Which is more acceptable may depend on the perceived relative importance of the behavior of the normal cost and of the actuarial liability.

3 Plan Amendments Under Projected Unit Credit

3.1 Projected Unit Credit

The projected unit credit method commonly is used, partly because the accounting bodies of both Canada and the United States require that it be used in calculating the pension expense to be entered in the employer’s financial statements (CICA,\(^1\) 1986, Section 3460.28;
Partly as a result, most Canadian and United States pension plans are valued for funding purposes using this method. The method is described under the names projected unit credit (Anderson, 1992, p. 152; Berin, 1989, p. 119), prorate accrued benefit (Trowbridge and Farr, 1976, p. 40), accrued benefit cost method (constant amount) (Winklevoss, 1977, p. 78), or projected accrued benefit cost method (McGill and Grubbs, 1989, p. 291).

Under the service prorate version of the projected unit credit method, the projected retirement age pension is allocated pro rata over years of pensionable service. Thus, \( B_i(x_j(t)) \) is based on pay projected to retirement and service accrued to age \( x_j \). The normal cost is the present value of the current year’s benefit allocation. The actuarial liability is the present value of the benefit allocated to the date of valuation at which the age is \( x_j \) nearest \( B_i(x_j) \). It is assumed that the date of valuation corresponds to the beginning of a plan year. Hence, the normal cost for the plan year for individual \( j \) is given by:

\[
PUCNC_i(t) = (B_i(x_j(t)) + 1 - B_i(x_j(t))) \frac{\bar{a}_y^{(12)}}{D_x(t)} \frac{D_y}{D_x(t)}
\]

and the actuarial liability by:

\[
PUCAL_i(t) = B_i(x_j(t)) \frac{\bar{a}_y^{(12)}}{D_x(t)} \frac{D_y}{D_x(t)}
\]

### 3.2 Plan Amendment

Prior to the plan amendment at \( t_A \) but at the attained age \( x_j(t_A) \) of individual \( j \) at the time of the valuation we have:

\[
B_p(x_j(t_A)) = r_0(x_j(t_A) - w_j)S_f^j.
\]

Again consider a nonretroactive increase at age \( x_j(t_A) \) of the benefit fraction from \( r_0 \) to \( r_1 \). Two possible methods of handling this situation are described next.

---

2 FASB refers to the Financial Accounting Standards Board.

3 SFAS refers to Statement of Financial Accounting Standards.
3.3 Variant 1: PUC Service Weighting

The plan document gives a definition of accrued benefit that may be used in obeying vesting legislation; this accrued benefit may be based on the salary at attained age $x_j(t)$.

Under variant 1 with $r_1 > r_0$, we assume instead that the benefit accrued up to age $x_j(t)$ is given by the fractional method:

\[ B_{A1}^j(x_j(t_A)) = \frac{(x_j(t) - w_j)}{(y_j - w_j)} \left[ r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A)) \right] S_f. \]  

The normal cost for the year following age $x_j(t)$, where $t \geq t_A$, would increase in the ratio:

\[ \frac{PUCN C_{A1}^j(x_j(t))}{PUCN C_p^j(x_j(t))} = \frac{r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A))}{r_0(y_j - w_j)}. \]  

This contrasts with the ratio $r_1/r_0$, which is more natural if one regards the benefit as accruing at a rate $r_0$, before the effective date of the amendment and at a rate $r_1$ afterward instead of using the fractional method.

The accrued liability under the fractional method at age $x_j(t_A)$ increases, because of the amendment, in the same ratio:

\[ \frac{PUC A_{A1}^j(x_j(t))}{PUC A_p^j(x_j(t))} = \frac{r_0(x_j(t_A) - w_j) + r_1(y_j - x_j(t_A))}{r_0(y_j - w_j)}. \]  

This increase in actuarial liability is somewhat counterintuitive in a situation where the benefit accrued up to age $x_j(t_A)$ can be regarded as being unchanged.

In the case $r_0 < r_1$, the normal cost and the actuarial liability are both decreased by the amendment.

3.4 Variant 2: PUC Accruals Weighting

Under variant 2, the benefit is assumed accrued at a rate $r_0$, for service before the amendment and $r_1$ for service afterward. It thus differs from variant 2 of the entry age method. Thus:

\[ B_{A2}^j(x_j(t_A)) = r_0(x_j(t_A) - w_j)S_f. \]
and the actuarial liability is unchanged:

\[
\frac{\text{PUC} A_{t}^j \left( x_j(t_A) \right)}{\text{PUC} A_{p}^j (x_j(t_A))} = 1. \tag{25}
\]

The normal cost at time \( t, t \geq t_A \), increases (or decreases if \( r_0 < r_1 \)) in the expected ratio because the variant 2 accrued benefit increases as:

\[
B_{A2}^j(x_j(t) + 1) = \left[ r_0(x_j(t_A) - w_j) + r_1(x_j(t) + 1 - x_j(t_A)) \right] S_j^j.
\tag{26}
\]

Hence

\[
\frac{\text{PUC} N C_{A2}^j(x_j(t))}{\text{PUC} N C_{p}^j(x_j(t))} = \frac{B_{A2}^j(x_j(t) + 1) - B_{A2}^j(x_j(t))}{B_{p}^j(x_j(t) + 1) - B_{p}^j(x_j(t))} = \frac{r_1}{r_0}. \tag{27}
\]

This variant gives results that might be expected by a client. In the United States, variant 2 usually is required for calculation of pension expense under SFAS 106 and SFAS 87 (Financial Accounting Standards Board, 1990, paragraph 40, footnote 8). Paragraph 40 of SFAS 87 states that “... pension benefits ordinarily should be based on the plan’s benefit formula to the extent that the formula states or implies an attribution.” Footnote 8 has “... benefit of 1 percent of final pay for each year of service up to 20 years and 1.5 percent of final pay for years of service in excess of 20 ... the attribution ... will not assign the same amount of pension benefit for each year of service.” If the plan document defines the benefit accrual on a fractional basis, as in equation (21), then variant 1 is acceptable.

In Canada, the requirements are less clear. The Canadian Institute of Chartered Accountants (1986, paragraph 3460.28) states “the cost of pension benefits ... should be determined using the projected benefit method prorated on services.”

The United States IRC Regulation Section 1.412(c)(3)-1(e)(3) discusses the allocation of projected benefits between past and future years. Example (5) of IRC Regulation Section 1.412(c)(3)-1(g) indicates variant 2 (PUC accruals weighting) to be the acceptable method for funding purposes when the plan document defines the accrued benefit according to equation (26) rather than according to the fractional equation (21). This variant also satisfies the zero gain
condition of IRC Regulation Section 1.412(c)(3)-1(c)(2), as is shown in Appendix C.

4 Conclusion

This paper discusses the use of the individual entry age normal and projected unit credit pension funding methods in the presence of a nonretroactive increase in the benefit accrual rate. In the case of both funding methods, it is recommended that the cost method be handled in such a way that the normal cost increases in the same proportion as the increase in the benefit accrual rate. Alternative methods are discussed, however, that may be more acceptable to some actuaries.

References


Canadian Institute of Chartered Accountants Handbook. (Canadian Institute of Chartered Accountants, 1986).


Appendix A—Entry Age Normal, Variant 2 (EAN Retroactive NC Mimic), Proof of Decrease of Actuarial Liability at Amendment

Let \( f(x) = \frac{sN_x - sN_y}{y - x} \)

and assume that \( sN_x \) is a decreasing function of \( x \). Then

\[
\begin{align*}
    f(x + 1) - f(x) &= \frac{sN_{x+1} - sN_y}{y - x - 1} - \frac{sN_x - sN_y}{y - x} \\
    &= \frac{(y - x)sN_{x+1} - (y - x - 1)(sN_{x+1} + sD_x) - sN_y}{(y - x - 1)(y - x)} \\
    &= \frac{sN_{x+1} - (y - x - 1)sD_x - sN_y}{(y - x - 1)(y - x)} \\
    &= \frac{(sD_{x+1} - sD_x) + (sD_{x+2} - sD_x) + \ldots + (sD_{y-1} - sD_x)}{(y - x - 1)(y - x)} \\
    &\leq 0 \quad (1.A)
\end{align*}
\]

with equality only if \( sD_x = sD_z \) for \( x + 1 \leq z \leq y - 1 \). Because \( w_j < x_j(t_A) \), equation (12) gives us

\[
EANA_{A2}(x_j(t_A)) = EANA_{A1}(x_j(t_A))
\]

\[
+ \frac{(r_1 - r_0)}{r_0} PVFB_p(x_j(t_A)) \times \frac{y_j - x_j(t_A)}{sN_{w_j} - sN_{y_j}} \\
\times \left[ \frac{sN_{w_j} - sN_{y_j}}{y_j - w_j} - \frac{sN_{x_j(t_A)} - sN_{y_j}}{y_j - x_j(t_A)} \right]
\geq EANA_{A1}(x_j(t_A))
\]

using the decreasing nature of \( f(x) \) from equation (1.A) and assuming \( r_1 > r_0 \).
Appendix B—Entry Age Normal, Variant 2 (EAN Retroactive NC Mimic), Proof of Zero Gain

The notation used is

\[ \begin{align*}
PL & = \text{For the whole plan;} \\
A & = \text{The set of actives (see Anderson, 1992, p. 9);} \\
G(t) & = \text{Gain in year } t \text{ to } t+1; \\
i & = \text{Valuation interest rate;} \\
F(t) & = \text{Fund value at time } t; \\
UAL_{PL}(t) & = \text{Unfunded at time } t, \text{ } UAL_{PL}(t) = AL_{PL}(t) - F(t) \\
C(t) & = \text{Actual contributions in the year } t \text{ to } t+1; \\
I_c(t) & = \text{Interest to time } t+1 \text{ at the assumed rate } i \text{ on the contributions } C(t). \\
\end{align*} \]

For simplicity, assume that the membership consists only of actives who will be below retirement age at the end of the year. Assume that the only benefit is on retirement. Use the standard formula for the gain (see, e.g., Anderson, 1992, p. 20). Assume, following IRC Regulation Section 1.412(c)(3)–1(c)(2), that “each actuarial assumption is exactly realized,” so that for example

\[ F(t + 1) - (C(t) + I_c(t) - F(t))(1 + i) = 0. \] (1.B)

Then the gain in a year \( t \), after the amendment, \( t \geq t_A \), is given by

\[ G(t) = \left( EAN_{UAL_A2(t)}^{PL} + EAN_{NC_A2(t)}^{PL} \right)(1 + i) - (C(t) + I_c(t)) \\
- EAN_{UAL_A2(t + 1)}^{PL} \\
= \left( EAN_{AL_A2(t)}^{PL} - F(t) + EAN_{NC_A2(t)}^{PL} \right)(1 + i) - (C(t) + I_c(t)) \\
- \left( EAN_{AL_A2(t + 1)}^{PL} - F(t + 1) \right) \\
= F(t + 1) - (C(t) + I_c(t)) - F(t)(1 + i) + \\
\left( EAN_{AL(t)}^{PL}_A2(t) + EAN_{NC}^{PL}_A2(t) \right)(1 + i) - EAN_{AL_A2(t + 1)}^{PL} \]

124
\[
\begin{align*}
&= \left( EANAL(t)_{A2}(t) + EANNC_{A2}(t) \right)(1 + i) - EANAL_{A2}(t + 1) \\
&= \sum_{A_t} \left( EANAL^j_{A2}(x_j(t)) + EANNC^j_{A2}(x_j(t)) \right)(1 + i) \\
&\quad - \sum_{A_{t+1}} EANAL^j_{A2}(x_j(t + 1)) \\
&= \sum_{A_t} \left[ PVFB^j_{A2}(x_j(t)) - PVF\left( EANNC^j_{A2}(x_j(t)) \right) \right] \left( 1 + i \right) \\
&\quad + EANNC^j_{A2}(x_j(t)) \left( 1 + i \right) \\
&\quad - \sum_{A_{t+1}} EAN \left[ PVFB^j_{A2}(x_j(t+1)) - PVF\left( EANNC^j_{A2}(x_j(t+1)) \right) \right] \\
&= \sum_{A_t} \left[ -PVF\left( EANNC^j_{A2}(x_j(t)) \right) + \left( EANNC^j_{A2}(x_j(t)) \right) \right] \left( 1 + i \right) \\
&\quad + \sum_{A_{t+1}} \left[ PVF\left( EANNC^j_{A2}(x_j(t+1)) \right) \right] \\
&= 0.
\end{align*}
\]

(2.B)

In the above has been used the assumption that decrements, which reduce \( A_t \) at time \( t \) to \( A_{t+1} \) at time \( t+1 \), give \( A_{t+1} \) as a proportion

\[
1 - q_{x_j(t)} = \frac{D_{x_j(t+1)}(1 + i)}{D_{x_j(t)}}
\]

of \( A_t \).
Appendix C—Projected Unit Credit, Variant 2 (PUC Accruals Weighting), Proof of Zero Gain

Use notation and assumptions as for Appendix B. Then the gain in the year starting at time \( t \) is given (Anderson, 1992, p. 13) using equation (1.B) by

\[
G(t) = \left( PUC_{UL}_{A2}(t) + PUC_{NC}_{A2}(t) \right)(1 + i)
\]

\[
- (C(t) + iC(t)) - PUC_{UL}_{A2}(t + 1)
\]

\[
= \left( PUC_{AL}_{A2}(t) + PUC_{NC}_{A2}(t) \right)(1 + i) - PUC_{AL}_{A2}(t + 1)
\]

\[
= \sum_{A_t} \left( PUC_{AL}_{A2}(x(t)) - PUC_{NC}_{A2}(x(t)) \right)(1 + i)
\]

\[- \sum_{A_{t+1}} PUC_{AL}_{A2}(x(t + 1))
\]

\[
= \sum_{A_t} \left[ r_0(x(t) - w) + r_1(x(t) - x(t)) + r_t \right] S(t) + i \frac{Dy_j}{Dx_j(t)} d_j^{12}
\]

\[- \sum_{A_{t+1}} \left[ r_0(x(t) - w) + r_1(x(t) + 1 - x(t)) \right] S(t) + i \frac{Dy_j}{Dx_j(t+1)} d_j^{12}
\]

\[= 0\]

where again the set \( A_t \) reduces after a year to \( A_{t+1} \) at the assumed proportion \((1 - q_{x(t)})\).
Estimating the Effect of Statutory Changes on Insured Casualty Losses Using Generalized Indicator Variables

Ruy A. Cardoso*

Abstract

Techniques for estimating future insured losses in casualty insurance typically assume consistency in the insurance environment over time. Statutory changes, however, can create sharp discontinuities in the loss-generating process, complicating the estimation of those losses. Using indicator variables and dummy variables allows for quantification of the effect of such discontinuities. Three examples from private passenger automobile insurance are presented to illustrate how these variables can be used.

Key words and phrases: dummy variables, linear regression, tort threshold, coverage stacking, coverage trigger, coverage limits

1 Introduction

Estimation of future insured losses in casualty insurance often is based on an examination of the past patterns of those losses over time. Usually a linear or exponential relationship between losses and time is postulated as a starting point. Under this traditional actuarial approach, a further implicit assumption is that the insured losses are generated by an underlying process that changes smoothly. Statutory changes, however, can create discontinuities in the loss-generating process that must be accounted for properly in estimating future losses. This paper explains and illustrates a simple method of accounting for such discontinuities after they have occurred. Specifically, the method uses generalized forms of the linear regression variables known as indicator (or dummy) variables. Section 2 describes the most common actuarial method of estimating future losses in the absence of such discontinuities, while Section 3 provides

* Ruy Cardoso, FCAS, MAAA, currently is vice president and chief actuary of the Automobile Insurers Bureau of Massachusetts where he represents the Massachusetts automobile insurance industry in regulatory hearings concerning insurance rates. Mr. Cardoso previously has held consulting and insurance company positions whose primary focus was the analysis of casualty insurance loss reserves. He graduated from the Massachusetts Institute of Technology in 1983 with an S.B. in management science.
some background on indicator variables. Section 4 provides several specific examples of the generalized indicator variable approach using Massachusetts private passenger automobile insurance data, and Section 5 briefly summarizes the advantages and disadvantages of this approach.

2 Traditional Estimation of Future Losses

For simplicity, the discussion below assumes that the quantity of interest in the estimation procedures is the pure premium or the average insured loss per unit of insurance exposure. For private passenger automobile insurance, the unit of insurance exposure is generally a car-year, i.e., a single car insured for one year. The two most common models used to estimate future pure premiums assume either a linear or an exponential relationship between pure premiums \( Y \) and time \( T \), as shown in equations (1) and (2):

\[
Y = a + bT
\]  \hspace{1cm} (1)

\[
Y = a e^{bT},
\]  \hspace{1cm} (2)

where \( a \) and \( b \) are constants (McClenahan, 1990). These two models often are based on economic indices rather than time and frequently include adjustments for autocorrelation (Cummins and Derrig, 1993). For simplicity in explaining the indicator variable approach, the remainder of this paper focuses on equation (1). Equation (2) sometimes is called log-linear because it can be transformed into equation (1) by taking logs. Once equation (2) is transformed, indicator variables also can be applied in a manner similar to that in equation (1). The interpretation of the quantities discussed below, however, would be different in the transformed case.

In practice, the time variable used in equation (1) is discrete, most often the accident year (the year in which the accident generating the loss occurred) associated with each loss. Further, the traditional method does not rely on individual losses. It works instead with aggregate pure premiums, in this case for each accident year. Thus, equation (1) simply says that pure premiums change by a constant dollar amount per year. Future pure premiums are estimated by assuming that the estimated annual change will continue into the future, although practicing actuaries often will modify the equation’s results if its underlying assumptions are too strict.

It is not necessary to attribute the estimated pure premium change to specific causes, although blind application of the model
may lead to unreasonable results, especially if the random component of the loss-generating process is high. Pure premiums are not a direct (causal) function of the time variable; time is intended as a proxy for the many unspecified factors that determine pure premiums. This lack of causal explanation, however, is common to many possible methods of estimating future pure premiums. For example, one may use a Box-Jenkins time series model (an approach widely used in non-actuarial settings) to relate the pure premium for a given accident year to pure premiums for past accident years and/or to past random errors, not to any underlying causal variables. The primary reason for using the time proxy is that, in practice, the number of available pure premium data points is usually too small to perform meaningful analyses of causal relationships (or, for that matter, Box-Jenkins analysis).

Whatever the underlying causal variables are, equation (1) implicitly assumes that they will behave smoothly over time. When there is a significant underlying change in the smoothness of the loss-generating process, the model is likely to produce poor estimates, making it necessary to deal with such discontinuities in some reasonable way. While the subjective adjustments frequently used in practice (for example, adjustment of data before the change to a postchange basis) may be appropriate in certain situations, the use of generalized indicator variables provides a more objective approach.

3 Background on Indicator Variables

An indicator random variable usually is defined with respect to the occurrence or non-occurrence of an event. Thus, if $A$ is an event and $I(A)$ is the indicator random variable of $A$, then

$$I(A) = \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise}. 
\end{cases}$$

In this paper, $A$ is assumed to be an event (a change in the environment) that affects the pure premium. (See Miller and Wichern (1977) for a brief discussion of indicator variables in linear regression analysis.) Incorporation of indicator variables into equation (1) produces the model shown in equation (3):

---

1 For a detailed description of the Box-Jenkins time series model and analysis, see Box and Jenkins (1970). For a brief introductory treatment, however, see Wheelwright and Makridakis (1985).
\[ Y = a + bT + \sum_{i=1}^{m} c_i I_i(A_i) \]

where \( m \) is the number of indicator variables used, \( c_i, i = 1, 2, \ldots, m \) are constants, and \( I_i \) is the indicator variable for the \( i^{th} \) change. Table 1 illustrates the results of such a model when \( m = 1, a = \$100, b = \$10, c_1 = \$5, I_1 = 0 \) for \( T \leq 3 \), and \( I_2 = 1 \) for \( T \geq 4 \). Here \( A \) is the event \( \{T \geq 4\} \).

**TABLE 1**

<table>
<thead>
<tr>
<th>( T )</th>
<th>( Y )</th>
<th>Change in Pure Premium</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$110</td>
<td>$10</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$120</td>
<td>$10</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>$130</td>
<td>$10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>$145</td>
<td>$15</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>$155</td>
<td>$10</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>$165</td>
<td>$10</td>
<td>1</td>
</tr>
</tbody>
</table>

Under equation (3), the indicator variable can be thought of as an on-off switch that reflects some change in the environment at and beyond \( T = 4 \). In a sense, a model using such a variable has one foot in the world of causal explanation.

It is not necessary for an indicator variable to be strictly zero-one, however. The terms *generalized indicator variables* and *dummy variables* are used interchangeably in this paper to reflect more general forms. Many changes in an environment are more analogous to a dimmer switch than to a simple on-off switch. That is, they occur gradually rather than all at once. McDowall, *et al.* (1980) describe the use of generalized indicator variables (or *intervention components* in their terminology) in Box-Jenkins time series analysis. The applications below will illustrate both the zero-one case and more general cases in the context of linear regressions against time, using statutory changes affecting private passenger automobile insurance as examples.

### 4 Specific Applications: Private Passenger Automobile Insurance

Permanent statutory changes in the insurance environment can have at least three effects on accident year pure premium data:

a) **Single step**, reflecting a change that is completely effective in a specified accident year and all subsequent accident years.
b) **Two step** (or, more generally, multiple step), reflecting a change that is partially effective in a specified accident year and completely effective in all subsequent accident years.

c) **Infinite step**, reflecting a change that is partially effective in a specified accident year and increasingly effective in all subsequent accident years, but never completely effective.

The first effect can be modeled using the simple zero-one indicator variable, the second using an indicator variable that takes values between zero and one, and the third using an indicator variable that takes values greater than one.

### 4.1 The Single Step Case

An example of a single step statutory change in private passenger automobile insurance is a change in the *tort threshold*, the level of injuries that must be sustained before a person injured in an automobile accident can sue for pain and suffering damages. Certain states have no restrictions on the right to sue (i.e., there is no tort threshold), while those states where a no-fault system exists have either a qualitative threshold (usually referred to as a *verbal threshold*) or a *monetary threshold* (usually measured by medical costs). In the state of Massachusetts the current tort threshold is a monetary one. That is, the medical costs of the injuries sustained in an accident must exceed a fixed dollar amount before a suit for pain and suffering can be filed. On January 1, 1989 this threshold was raised from $500 to $2,000 for all accidents occurring on or after January 1, 1989. It follows that:

\[
I = \begin{cases} 
1 & \text{if accident year } \geq 1989 \\
0 & \text{otherwise.}
\end{cases}
\]

Table 2 displays the accident year pure premiums for the bodily injury liability (BIL) coverage for the accident years 1984-1992. (Losses are limited to basic limits and developed to ultimate values.) Figure 1 displays the values in graphical form.
As both Table 2 and Figure 1 show, pure premiums rose steadily over the accident years 1984-1992 except in accident year 1989, the year in which the tort threshold was raised. As had been expected, raising the threshold reduces the pure premiums for the bodily injury liability coverage. An estimate of how much the pure premiums were reduced can be obtained using linear regression of the pure premiums against both the accident years and an indicator variable that is assigned the value of zero in accident years 1984-1988 and the value
of one in accident years 1989-1992. The regression results are shown in equation (4).

\[ PP = 57.39 + 9.83 \times T - 12.19 \times I \]  \hspace{1cm} (4)

where \( PP \) denotes pure premium, \( T \) denotes accident year (with 1984 considered year 1), and \( I \) denotes the indicator variable. Fitted values according to this equation also are displayed in Figure 1. The interpretation of equation (4) is that pure premiums are rising at $9.83 per year and that the change in the tort threshold reduces pure premiums by $12.19 from what they otherwise would have been (although the t-statistic for the coefficient of the indicator variable is not significantly different from zero under the usual significance levels). Future pure premiums in the presence of the higher tort threshold can be estimated using the above equation and holding the indicator variable at its postchange value of one. Naturally, the use of fewer data points will result in different estimates. This model's residuals indicate serial correlation of the errors, although the serial correlation might disappear if the infinite step model described in Section 4.3 were used. Analysis of residuals, however, is not a topic for this paper. Equation (4) simply serves to show how a zero-one indicator variable can be applied.

4.2 The Two-Step Case

Because private passenger automobile insurance policies are written throughout a given calendar year, the policy that covers an accident occurring in a particular accident year may have been written in that year or in the prior year. A change in the terms of the policy, therefore, will not affect all accidents occurring in a given year, only those covered by policies written after the change. In other words, a policy change will have only a partial effect on the accident year in which the change is made.

At the same time the tort threshold was raised in Massachusetts, another pair of statutory changes led to just this effect. A stacking provision (which determines whether policy limits from multiple policies in the same household can be combined) and a trigger provision (which determines the conditions under which coverage applies) were both modified in a way that was expected to reduce pure premiums. These modifications only applied to uninsured/underinsured motorists (UM/UIM) coverages, which pay for injuries in which a driver has insufficient bodily injury liability insurance (if any) to cover an insurance claim arising from an accident he or she caused.
Prior to January 1, 1989, households with more than one UM/UIM policy, under certain circumstances, could combine (stack) the limits of all of those policies to cover a single accident, in effect multiplying the limit on each policy by the number of policies in the household (if the limits were the same for each policy). This ability to combine limits was removed for policies written on or after January 1, 1989, reducing aggregate losses paid from what they otherwise would have been.

The change in the trigger provision works as follows. Losses paid under the UM/UIM coverages were unaffected by the limits of an at-fault driver’s bodily injury liability insurance until January 1, 1989. Policies written on or after that date, however, only pay losses up to the difference in limits between the UM/UIM coverage, and the at-fault driver’s bodily injury liability limits. (That is, an additional constraint must be satisfied before the coverage is triggered.)

Because of the effective date of these changes, they were only partially effective in accident year 1989 but completely effective in all subsequent accident years. Based on the distribution of inception dates for policies written in Massachusetts, about 65 percent of the accidents occurring in accident year 1989 were covered by the modified policy. Table 3 displays the accident year pure premiums for the UM/UIM coverages for the accident years 1984-1992. (Losses are limited to basic limits and developed to ultimate values.) Figure 2 displays the values in graphical form.

<table>
<thead>
<tr>
<th>Acc Year</th>
<th>T</th>
<th>UM/UIM Pure Premium</th>
<th>Change in Pure Premium</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1</td>
<td>$18.91</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>1985</td>
<td>2</td>
<td>$23.83</td>
<td>$4.92</td>
<td>0</td>
</tr>
<tr>
<td>1986</td>
<td>3</td>
<td>$26.91</td>
<td>$3.08</td>
<td>0</td>
</tr>
<tr>
<td>1987</td>
<td>4</td>
<td>$29.40</td>
<td>$2.49</td>
<td>0</td>
</tr>
<tr>
<td>1988</td>
<td>5</td>
<td>$33.56</td>
<td>$4.16</td>
<td>0</td>
</tr>
<tr>
<td>1989</td>
<td>6</td>
<td>$20.91</td>
<td>($12.65)</td>
<td>0.65</td>
</tr>
<tr>
<td>1990</td>
<td>7</td>
<td>$17.50</td>
<td>($3.41)</td>
<td>1</td>
</tr>
<tr>
<td>1991</td>
<td>8</td>
<td>$19.27</td>
<td>$1.77</td>
<td>1</td>
</tr>
<tr>
<td>1992</td>
<td>9</td>
<td>$20.20</td>
<td>$0.93</td>
<td>1</td>
</tr>
</tbody>
</table>

2 The losses paid under the UM/UIM coverages also should have been affected by the change in the tort threshold, but to a far lesser degree than they were affected by the stacking and trigger changes.
As both Table 3 and Figure 2 show, pure premiums rose steadily over the accident years 1984-1992 except in accident years 1989 and 1990, the two years over which the modified stacking and trigger provisions became effective. As expected, the two changes reduce the pure premiums for the UM/UIM coverages. An estimate of how much the pure premiums were reduced can be obtained using linear regression of the pure premiums against both the accident years and a generalized indicator variable (dummy variable) $I$ that is assigned the value of zero in accident years 1984-1988, the value of 0.65 in accident year 1989, and the value of one in accident years 1990-1992, i.e.,

$$I = \begin{cases} 
0 & \text{if } T = 1, 2, 3, 4, 5 \\
0.65 & \text{if } T = 6 \\
1 & \text{if } T \geq 7.
\end{cases}$$

The regression results are shown in equation (5):

$$PP = 17.14 + 3.13 \times T - 23.17 \times I$$

(5)
where $PP$ denotes pure premium, $T$ denotes accident year (with 1984 considered year 1), and $I$ denotes the dummy variable. Fitted values according to this equation also are displayed on Figure 2. The interpretation of equation (5) is that pure premiums are rising at $3.13 per year and that the stacking and trigger modifications reduce pure premiums by $23.17 from what they otherwise would have been. As the indicator variable for accident year 1989 is 0.65, the reduction in that year was not the full value of $23.17$, however; it was instead a partial value of $(0.65 \times 23.17)$ or $15.06$. Future pure premiums under the modified stacking and trigger provisions can be estimated using the above equation and holding $I=1$. Again, the use of fewer data points will result in different estimates. Despite the two step nature of the discontinuity in this case, the functional form of the equation is the same as that of equation (4). Both are simply special cases of the general equation (3).

It is important to note, however, that the two step case described above also could be modeled using two zero-one indicator variables, the first changing to one in 1989 and the second changing to one in 1990. While the results of such a model would be similar to the results produced by equation (5) (due to the close fit), they would come at the cost of a degree of freedom and a less apparent model structure. It is easy to grasp the concept of a partial effect by seeing a generalized indicator variable with a value of 0.65, and it is clear in this instance that the 0.65 value has an objective basis rather than one that only pretends not to steal a degree of freedom.

4.3 The Infinite-Step Case

Certain changes in the insurance environment not only shift the relationship between pure premiums and time but also change the slope of the relationship. This type of effect can be modeled using two generalized indicator variables, the first the usual zero-one type and the second comprising a series of infinitely increasing values. The particular change in Massachusetts that can be modeled this way occurred at the same time as the change in the tort threshold and was effective for all accidents occurring in accident year 1989 and subsequent years (despite contrary policy language). Specifically, the coverage limit of the personal injury protection (PIP) coverage increased from $2,000 to $8,000 on January 1, 1989. This coverage pays for injuries regardless of fault and therefore also is known as no-fault coverage.

Because many of the claims paid under the PIP coverage reached the $2,000 limit in the years before the limit was increased, claim
cost inflation only could affect a subset of all claims. Increasing the limit to $8,000, however, allows those claims previously constrained by the limit to reflect the effects of claim cost inflation, in turn allowing the aggregate pure premiums for the PIP coverage to reflect inflation more completely and thus increase more quickly (i.e., with a greater slope). While it is possible that increasing the tort threshold also may have a slope-changing effect on BIL coverage (see Section 4.1), the PIP limit change serves as a much clearer illustration.

If we denote the slope of the pure premium line under the $2,000 limit as \( b \), the size of the discontinuity created by the statutory change as \( c_1 \), and the slope of the pure premium line under the $8,000 limit as \( c_2 \) (where \( c_2 \) is expected to be greater than \( b \)), then pure premiums over time can be modeled as follows:

\[
PP = \begin{cases} 
  a + bT & \text{for } T \leq 5 \\
  a + bT + c_1 & \text{for } T = 6 \\
  a + 6b + c_1 + c_2(T-6) & \text{for } T \geq 7.
\end{cases}
\] (6A)

While this is a natural way to model the PIP pure premiums over time, equation (6A) does not fit into the general equation (3). In order to transform equation (6A) into a specific instance of equation (3), it is necessary to redefine \( c_2 \) as the difference between the post-1989 slope and the pre-1989 slope (where the difference is expected to be positive) and adopt the following pair of generalized indicator variables:

\[
I_1 = \begin{cases} 
  0 & \text{if } T \leq 5 \\
  1 & \text{if } T \geq 6
\end{cases}
\]

and

\[
I_2 = \begin{cases} 
  0 & \text{if } T \leq 6 \\
  (T-6) & \text{if } T \geq 7.
\end{cases}
\]

Equation (6A) can be recast as:
\[ PP = \begin{cases} a + bT & \text{for } T \leq 5 \\ a + bT + c_1 & \text{for } T = 6 \\ a + bT + c_1 + c_2(T-6) & \text{for } T \geq 7, \end{cases} \tag{6B} \]

with equation (6B) being a specific instance of equation (3):

\[ PP = a + bT + c_1 l_1 + c_2 l_2. \tag{6C} \]

Table 4 below displays the accident year pure premiums for the PIP coverage for the accident years 1984 to 1992 (again developed to ultimate values); Figure 3 displays the values in graphical form.

<table>
<thead>
<tr>
<th>Acc Year</th>
<th>T</th>
<th>PIP Pure Premium</th>
<th>Change in Pure Premium</th>
<th>Indicator #1</th>
<th>Indicator #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1</td>
<td>$12.98</td>
<td>NA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1985</td>
<td>2</td>
<td>$14.97</td>
<td>$1.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1986</td>
<td>3</td>
<td>$15.92</td>
<td>$0.95</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1987</td>
<td>4</td>
<td>$17.61</td>
<td>$1.69</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1988</td>
<td>5</td>
<td>$19.63</td>
<td>$2.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1989</td>
<td>6</td>
<td>$36.03</td>
<td>$16.40</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1990</td>
<td>7</td>
<td>$39.81</td>
<td>$3.78</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1991</td>
<td>8</td>
<td>$43.39</td>
<td>$3.58</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1992</td>
<td>9</td>
<td>$48.33</td>
<td>$4.94</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

As both Table 4 and Figure 3 show, pure premiums rose steadily over accident years 1984 through 1988, jumped sharply at accident year 1989, and rose more steeply over accident years 1990 through 1992 (as expected). An estimate of how much the pure premium line was shifted and steepened because of the change in limit can be obtained using linear regression of the pure premiums against both the accident years and the two indicator variables displayed in Table 4 above. The regression results are shown in equation (7).

\[ PP = 11.44 + 1.59 \times T + 14.81 \times l_1 + 2.45 \times l_2 \tag{7} \]
where \( PP \) again denotes pure premium, \( T \) denotes accident year, and \( I_1 \) and \( I_2 \) denote the indicator variables. Fitted values according to this equation are displayed on Figure 3. The interpretation of equation (7) is that pure premiums were rising at $1.59 per year, increased $14.81 as a result of the change in the PIP coverage limit (because \( I_1 = 1 \) in 1989), and now are rising at $4.04 per year, where $4.04 equals the prechange slope of $1.59 plus the postchange increment of $2.45 (the coefficient of \( I_2 \)). Future pure premiums under the $8,000 PIP coverage limit can be estimated by using the above equation, holding \( I_1 \) constant at its value of one and moving \( I_2 \) up one for each year beyond accident year 1989.

Relative to the single step and two step cases, this case has cost another degree of freedom. But in this situation an additional quantity is being estimated, specifically the postchange slope, making the cost an appropriate one to pay. Further, the model structure is reasonably apparent. While other approaches could be used to model the infinite step case, the one used here strikes the best balance between clarity and degrees of freedom.

5 Summary of the Approach

As illustrated above, generalized indicator variables can be used to model a variety of different time series discontinuities in private passenger automobile insurance. While the examples above have
been restricted to permanent statutory changes, the approach can be extended easily to temporary changes as well as to other lines of insurance. This flexibility is a key advantage of the approach, as is its ability to let the data speak for themselves. Alternative approaches, such as adjusting all data to a postdiscontinuity basis, can work in the single step case above, but such an alternative is likely to be more subjective than the generalized indicator variable approach.

On the other hand, a too-complicated set of indicator variables could be used to mask the occasional tendency to force a preordained conclusion. Further, the use of multiple indicator variables easily could lead to overfitting, especially in the common situation where only a small number of data points is available. Such pitfalls should not blind the actuary to the usefulness of the generalized indicator variable approach. As with any model-building exercise, the value of indicator variables as a tool will rise with the care taken in using them.

References


Ruy A. Cardoso
Automobile Insurers Bureau of Massachusetts
101 Arch Street
Boston, Massachusetts 02110
Life Insurance Applications of Recursive Formulas

L. Timothy Giles

Abstract

This paper discusses several practical applications of recursive formulas:

a) Traditional whole life—As an introduction, the well-known relationship between successive terminal reserves is reviewed. Recursive formulas are developed to calculate the reserves and the premiums;
b) Universal life—Recursive formulas are used both for the calculation of target premiums and reserves. Consideration is given to the TEFRA corridor;
c) Paid-up rider—A participating single premium rider that provides a level death benefit can be devised using an inherent one year term benefit. Recursive functions are used to determine the premium that precisely matures the rider.

Because the APL programming language is particularly amenable to recursive formulas, a few sample APL programs are provided.

Key words: APL, TEFRA corridor, universal life, paid-up rider

1 Introduction

A recursive formula is one where the current result is generated from previous results once the starting values are given. Essentially, a recursive formula is a difference equation with known starting values. Some formulas are special linear difference equations that, when added, condense to the first and last values of the recursion. Thus, if the starting value is known (usually zero), the ending value (usually the maturity amount) and all the intermediate values can be derived easily.

* L. Timothy Giles, FSA, MAAA is an actuary at Farm Family Life Insurance Company, Albany, New York. He began his actuarial career as a summer student at New York Life Insurance Company while majoring in mathematics at Holy Cross College. He has worked at six other life insurance companies during his thirty year career.

1 This paper is an extension of the author's paper entitled “The Practical Use of Recursive Functions” that appeared in ARCH, 1993.
Actuaries are familiar with such formulas, especially in the area of reserve and asset share calculations; see, for example, Jordan (1991) and Bowers et al. (1986). Shiu (1987) and Seah and Shiu (1987) elegantly present recursive formulas in their discussions of papers by Berin and Lofgren (1987) and Eckley (1987), respectively.

My interest in recursive formulas began in 1985 with the problem of finding the target premium for universal life policies. What level deposit (along with credited interest and mortality and expense charges) would mature the policy exactly? At that time I used trial and error to obtain the target premiums. A much better method, however, later was published by Eckley (1987). What these pioneers discovered is that by defining a transformed mortality rate $Q' = Q/(Q + 1)$, the traditional commutation functions could be used (with $Q'$ replacing $Q$) to calculate premiums and account values directly. These modified commutation functions are called transformation functions.

When the accumulation formula changes due to the TEFRA3 corridor (if the cash value becomes too high in relation to the death benefit, a higher minimum death benefit is invoked), however, a different formula for $Q'$ must be used. One, theoretically, could switch functions at the duration of change. I believe, however, that recursive formulas offer a better solution. Recursive formulas also can be used to perform the intricate calculation of paid-up riders that fund a benefit with a combination of one year term and paid-up insurance.

2 Introduction to the Method

Following Shiu (1987), define a first order linear recursive (difference) equation as

$$x_{k+1} = a_k x_k + b_k \quad k = 0, 1, 2, \ldots$$

(1)

where $x_0$, $\{a_k\}$, and $\{b_k\}$ are known. Let

$$s_{k+1} = a_0 a_1 \ldots a_k = \prod_{r=0}^{k} a_r$$

(2)

with $s_0 = 1$.

Dividing both sides of equation (1) by $s_{k+1}$ yields

---

2 I am also familiar with a similar (though unpublished) work by Wesley C. Green. Mr. Green currently is director of Actuarial Systems and MIS at Phoenix Home Life Mutual Insurance Company in Hartford, CT.

\[
\frac{x_{k+1}}{s_{k+1}} = \frac{x_k}{s_k} + \frac{b_k}{s_{k+1}}
\]
i.e.,
\[
\Delta \left( \frac{x_k}{s_k} \right) = \frac{b_k}{s_{k+1}}.
\]
Summing equation (3) from \(k=0\) to \(k=n-1\) yields
\[
x_n = x_0 + \sum_{k=0}^{n-1} \frac{b_k}{s_{k+1}} \times s_n.
\]
Thus, once the starting value \(x_0\) is determined, the \(n^{th}\) term can be obtained directly without explicitly computing the intermediate terms.

3 Traditional Whole Life

Here is the procedure for deriving a recursive formula for a traditional whole life reserve with a death benefit of 1:

a) Establish the succession rule.
You must know precisely the mathematical relationship between the reserve at \(t\) and that at \(t+1\). For example, for traditional whole life with face value \(t\), it is well known (see, for example, Bowers et al., Chapter 7, Section 8) that the succession rule for the reserves is:

\[
(iV + P) \times (1+i) = t+1V \times (1-Q_t) + Q_t
\]

where \(P\) is the net level premium, \(iV\) is the net premium terminal reserve at time \(t\), \(Q_t\) is the valuation mortality rate at age \(x+t\) (\(x\) is the issue age), and \(i\) is the valuation interest rate.

b) Cast the succession rule into linear form.
Put equation (5) in linear form (as in equation (1)),

\[
t+1V = iV \times \frac{(1+i)}{(1-Q_t)} + P \times \frac{(1+i)}{(1-Q_t)} - \frac{Q_t}{(1-Q_t)}.
\]

c) Compute the compounding element from issue to maturity.
The compounding element \(F_k\), as in equation (2), gives
\[ F_{k+1} = \prod_{j=0}^{k} \left[ \frac{(1+i)}{(1-Q_j)} \right]. \]

Notice that \( 1/F_k \) is a discount factor, i.e.,

\[ \frac{1}{F_k} = (1+i)^k P_x = \frac{D_{x+k}}{D_x}. \]

d) Divide both sides of the succession rule by \( F_{t+1} \) and place the result in finite difference form as in equation (3).

\[
\Delta \left( \frac{V}{F_t} \right) = \left[ \frac{P \times ((1+i)/(1-Q_t))}{F_{t+1}} \right] - \left[ \frac{Q_t/(1-Q_t)}{F_{t+1}} \right]
\]

\[ = \frac{P}{F_t} - \frac{Q_t/(1-Q_t)}{F_{t+1}}. \]

e) Sum both sides from issue to maturity.
Let \( m \) be the number of years from issue to maturity. Note that \( mV = 1 \) and \( 0V = 0 \).

\[
\frac{V_m}{F_m} = \sum_{k=0}^{m-1} \left[ \frac{P}{F_k} - \frac{Q_k/(1-Q_k)}{F_{k+1}} \right]. \tag{7}
\]

f) Solve for \( P \).
From equation (7),

\[
P \times \sum_{t=0}^{m-1} \frac{1}{F_t} = \frac{1}{F_m} + \sum_{t=0}^{m-1} \frac{vQ_t}{F_t} \tag{8}
\]

where \( v = (1+i)^{-1} \). Note that equation (8) can be rearranged as follows

\[
\frac{mV}{F_m} - \frac{0V}{F_0} = P \times \sum_{t=0}^{m-1} \frac{1}{F_t} - \sum_{t=0}^{m-1} \frac{vQ_t}{F_t}
\]

which essentially states that
Maturity value – Issue value = Premiums – Claims,

with all terms discounted to issue.

g) **Generate the intermediate values from the succession rule and the premium.**

From equation (4), it follows that

\[ tV = F_t \times \sum_{k=0}^{t-1} \frac{P - v Q_k}{F_k} \cdot \]

Notice that we have discounted all terms to issue, then accumulated.

## 4 Universal Life

This demonstration assumes a level death benefit of 1. The actuarial starting point for universal life mathematics is equation (5), the formula connecting successive terminal reserves for conventional whole life insurance:

\[ (tV + P)(1+i) = t+1V + Q_t \times (1 - t+1V) \cdot \]

Here \( 1 - t+1V \) is called the *net amount at risk*, and \( Q_t \times (1 - t+1V) \) is called the *mortality charge*.

Three sources of difficulties have to be overcome:

a) The mortality charge, being an expense, is payable at the beginning of the period, whereas the whole life reserve formula assumes deaths occur at the end of the year.

b) The net amount at risk is defined contractually. The traditional definition cannot be used because the end of the year reserve (or cash value) is not known at the time the mortality charge is due, which is at the beginning of the time period. A common approximation\(^4\) is to use

\[ t+1V \approx (tV + P)(1+i). \quad (9) \]

\(^4\) An alternative approximation is to include the mortality charge in the approximation, i.e.,

\[ t+1V \approx (tV + P)(1+i) - Q_t \times (v - (V + P)) \cdot \]
Discounting the approximate mortality charge at interest results in \( Q_t \times (v - (iV + P)) \) being the current mortality charge. Note that \( i \) is the guaranteed rate because the current rate is unstable.

c) The crediting and charging is done monthly, not annually as in the whole life reserve.

Monthly interest and mortality charges can be used directly or an algorithmic adjustment to annual processing can shorten the computation time slightly. Eckley explains this, noting that the mortality charge and the interest credit are constant for all 12 months of a policy year, at least in sales illustrations. The interested reader is referred to Eckley's paper for the formulas. The recursive formulas will work if monthly mortality is used, although the vector will be 12 times as large. A monthly interest rate is installed easily.

An accurate target premium also must include expense charges. The succession rule must be based upon the actual administrative processing (typically done by a mainframe computer). If that processing is in place, the actuary has to try to mimic the routines used on the mainframe instead of establishing theory.5

There is, of course, a different succession rule when the TEFRA corridor is in effect. To qualify as a life insurance product in the United States and receive the attendant tax benefits, there must be a minimum relationship between the death benefit and the cash surrender value. This is called the cash value (or TEFRA) corridor and is 250 percent for attained ages 40 and under and gradually reduces to 100 percent at age 95. There is no sound actuarial basis for these ratios; they are simply products of the U.S. Congress.

The most common industry response to this requirement is to include a contractual benefit setting the death benefit equal to the minimum of the face amount or the TEFRA corridor multiplied by the cash surrender value. This response in turn generates a complex NAIC6 reserve mandate that in effect requires the immediate funding of any projected triggering of the corridor. If the cash surrender value at the valuation date is such that its accumulation along with future guaranteed maturity premiums will cause the death benefit to exceed

5 An actuary, at an earlier time, may have established the processing rule. There are details (particularly when the minimum death benefit is in effect) such as the definition of the net amount of risk that are subject to choice. The practicing actuary then has to mimic the succession rule that is in place.

6 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.
the face amount, the present value of the excess is added to the
reserve. In some cases the calculation has to jump between the two
rules, but usually not more than once.

The succession rule for universal life (using the approximation to
t_{t+1}V given in equation (9)) is

\[ t_{t+1}V = (tV + P) \times (1 + i) - Q_t \times (v - tV - P) \times (1 + i). \]

For the TEFRA corridor, the death benefit is the cash value times 
\((1 + \Theta_t)\), where \(\Theta_t \geq 0\). The cash value at the end of period \(t+1\), where
death is assumed to occur, is \(t_{t+1}CV\), where \(t_{t+1}CV = (tV + P) \times (1 + i)\).
Thus, the net amount at risk at that time is

\[ (1+\Theta)(tV + P) (1 + i) - (tV + P)(1 + i) = \Theta_t \times (tV + P)(1 + i). \]

The annual mortality charge is paid at the beginning of the year.
Hence:

\[ t_{t+1}V = (tV + P) \times (1 + i) - Q_t \times (tV + P) \times \Theta_t. \]

5 Paid-Up Insurance Rider

The conventional paid-up addition is a single premium participating product. Each dividend then buys another portion of paid-up insurance, which is itself participating. The effect is an increasing amount of paid-up insurance.

A variation of this effect is a level death benefit participating single premium whole life policy. The facilitating device is to divide the death benefit into a one year term portion and a paid-up portion. The dividend first pays for the next year's term insurance with the same death benefit; any remaining dividend buys a paid-up participating whole life benefit. If the dividend is not sufficient to pay for the term portion, some of the paid-up could be surrendered and the face value of the term increased. If reduced dividends continued, there will be a time when the death benefit would have to be decreased. The initial premium rate is designed to pay for both the term portion and the paid-up portion, thus avoiding paying for the term in arrears.

Let \(P = \) the single premium. It buys a combination of paid-up, \(PU\), and a one year term insurance determined from:

\[ PU \times A_x + (1 - PU) \times c_x = P \]
\[ P = (A_x - c_x) \left[ \frac{1}{F_m} - \sum_{t=0}^{m-1} \frac{(\text{DIV}_{t+1} - c_{x+t+1})}{(A_{x+t+1} - c_{x+t+1}) F_{t+1}} \right]. \]

The result is the exact premium that will mature the policy if dividends are paid as projected; a daunting task by trial and error.

There are constraints that should be noted. If the paid-up amount were to exceed 1, future premiums would not be accepted. Also, negative paid-up amounts are not admissible.

6 Recursion Formulas in APL

APL programming language is well-suited to recursive formulas, as partial products are generated easily. A good APL technique for accumulating nonlevel payments is to discount all of the payments to the present, then accumulate this lump sum. The same process is used with recursive formulas.

Refer to the first practical example of traditional whole life,

\[ F_t = x \times (1 + i) \div (1 - \text{Q}_t) \text{ and } F_m = 1 \sum F_t. \]

In APL,

\[ P = (((1 + F_m) \div (\text{Q}_t + (1 - \text{Q}_t)) + F_t) + (-1 \downarrow 1, V1_t)) \]

The nonlevel payment accumulation technique described above can be used for intermediate values of the traditional whole life reserve:

\[ _t V = F_t \times (+\text{P} \div -1 \downarrow 1, F_t) - \text{+Q}_t + (1 - \text{Q}_t) + F_t \]

7 Conclusion

For the past several decades, commutation functions have served actuaries well. The more complex products of today, however, call for new techniques. Recursive formulas may be the answer. They serve best where the calculation involves a trajectory to a target. Even whole life can be viewed as finding a premium to mature the policy. The intermediate values, the reserves, follow easily. The tradition
of valuations made with a table of stored reserve factors can be improved. The reserves can be calculated on an as needed basis (which is the way universal life reserves must be calculated). Commutation functions also could be used to calculate reserves as needed. Recursive formulas provide a second way. I hope future actuaries are as comfortable with recursive formulas as present actuaries are with commutation functions.

References


L. Timothy Giles  
*Farm Family Life Insurance Company*  
P.O. Box 656  
*Albany, New York 12210*