1996


Colin Ramsay, Editor

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

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

Part of the [Accounting Commons](https://digitalcommons.unl.edu/accounting_commons), [Business Administration, Management, and Operations Commons](https://digitalcommons.unl.edu/businessadministration_commons), [Corporate Finance Commons](https://digitalcommons.unl.edu/corporatefinance_commons), [Finance and Financial Management Commons](https://digitalcommons.unl.edu/financeandfinancialmanagement_commons), [Insurance Commons](https://digitalcommons.unl.edu/insurance_commons), and the [Management Sciences and Quantitative Methods Commons](https://digitalcommons.unl.edu/managementcommonscq)


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

Methodologies for Determining Reserve Liabilities in the Workers Compensation High Deductible Program
Jerome J. Siewert ................................................................. 179

Third Party Administrator (TPA) Service Pricing and Incentive Contracts
Hou-Wen Jeng ................................................................. 209

Annuity Choices for Pensioners
M. Zaki Khoransee ...................................................... 229

Pension Funding by Normal Costs or Amortization of Unfunded Liabilities
Keith P. Sharp ............................................................. 257

What We Say in the NAIC Annual Statement Blank Actuarial Opinion
Kenneth W. Faig, Jr. ..................................................... 275

Constrained Forecasting of the Number of IBNR Claims
Louis G. Doray ............................................................. 287
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 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 to the Journal first is screened by the editor for suitability. If it is deemed suitable, copies are sent to at several independent referees. The name of the author(s) of the paper under consideration is usually anonymous to the referees, and the identities of referees are never 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 decisions: (1) accept subject to minor revisions, (2) accept subject to major revisions, or (3) reject.

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

See back cover for instructions to authors.
INSTRUCTIONS TO AUTHORS

Submissions: Send five copies of the manuscript to: Colin M. Ramsay, Journal of Actuarial Practice, P.O. Box 22098, Lincoln, NE 68542-2098, USA. Submission of an article implies it is not being submitted for publication elsewhere and has not been published elsewhere.

Manuscript Preparation: The text should be typed double-spaced on one side of each page. The top, bottom, left, and right margins should be 1 inch. The first page should include the title, author(s), institutional affiliation, address, and telephone and fax numbers. The second page should include only the title, the abstract (summary), and key words. The abstract must consist of no more than 150 words. Do not include references, footnotes, or abbreviations in the abstract. A list of no more than six major key words and phrases should be placed below the abstract. Each key word or phrase must express an important idea in the paper and must not appear in the title. Authors must check their manuscripts for clarity, grammar, spelling, punctuation, and consistency of references.

Literature Review: Authors must review the existing literature in order to place their work in a proper context, i.e., explain the relationship of the paper to other related research in the field.

Equations: Number consecutively only those equations referenced in the text. Indent equations and place equation numbers in parentheses at the right margin. Type equations with one space before and after mathematical function signs except in subscripts and superscripts. Type the main components, the equal sign, and the fraction bar on the main line. In the text, refer to equations in the following manner: "... equation (15) ..."

Tables: Tables must be integrated in the text. Each table must occur just before or just after the first reference to it in the text. Center the word "Table" followed by an Arabic numeral above the body of the table, e.g., Table 5. Separate headings in a table from the title of the table and from the body of the table with solid lines. Use a solid line to end the table. In the text, refer to tables as "... in Table 3 ...." Tables should be self-contained, i.e., readers must be able to interpret the table without looking in the text for variable definitions.

Figures: Authors must provide figures in camera-ready form of a professional quality; see the section "Accepted Papers" below. (Photocopies are acceptable initially.) Like tables, refer to a specific figure in the text as "_ in Figure 8 _." When labeling figures, capitalize the first letter in the word and number with Arabic numerals (e.g., Figure 1). In figure titles, capitalize the first letter of each word. Figures must be integrated into the text.

References: References must be placed at the end of the text, but before any appendices. List references alphabetically by author's last name. Include only those references cited in the text. Examples:


Accepted Papers: As the Journal is published using the LaTeX2e typesetting package, the final revision of articles accepted for publication must be submitted on a 3½ inch diskette containing the article (including tables and figures) saved in a format that is easily converted into LaTeX. (Authors who do not have access to LaTeX must contact the editor for information of submitting articles in Microsoft Word or WordPerfect.) In addition, figures must be saved separately on the diskette as graphics files in one of the following formats (with extension): *.TIF, *.EPS, *.PCX.
Methodologies for Determining Reserve Liabilities in the Workers Compensation High Deductible Program

Jerome J. Siewert*

Abstract†

In this paper I describe several approaches for estimating liabilities under a high deductible program, including a proposal for a more sophisticated approach relying upon a loss distribution model. The discussion addresses several related issues dealing with deductible size and mix, absence of long-term histories, and the determination of consistent loss development factors among deductible limits. In addition, I propose several approaches for estimating aggregate loss limit charges, if any, and the asset value for associated servicing revenue.

Key words and phrases: loss ratio, excess loss, ultimate loss, IBNR, development factors, inverse power curve

1 Introduction

With the advent of the workers compensation high deductible program in the early 1990s, actuarial efforts focused principally on pricing issues. Insurers initially developed this program to:

*Jerome J. Siewert, F.C.A.S., M.A.A.A., is an assistant vice president and actuary at the Wausau Insurance Companies. He manages the reserving unit of the actuarial department. Mr. Siewert’s previous experience includes managing several pricing units and serving on various Casualty Actuarial Society ratemaking committees. Prior to joining Wausau, he spent four years as a secondary school mathematics teacher.

Mr. Siewert’s address is: Wausau Insurance Companies, PO Box 8017, Wausau WI 54402, USA. Internet address: jsiewert@e-mail.com

†This paper is based on a previous paper entitled “A Model for Reserving for Workers Compensation High Deductibles” that appeared in The Casualty Actuarial Society Forum (Summer 1996), a nonrefereed publication of the Casualty Actuarial Society.
• Achieve price flexibility while passing additional risk to larger insureds in what was considered at that time an unprofitable line of business;

• Ameliorate onerous residual market charges and premium taxes in some states;

• Realize cash flow advantages similar to those of the closely related product, the paid loss retro;

• Provide insureds with another vehicle to control losses while protecting them against random, large losses; and

• Allow self-insurance without submitting insureds to sometimes demanding state requirements.

As the program matured, insurers' focus shifted to the liability side. Questions now are being asked about what losses are emerging and what they ultimately will cost insurers. For the actuary, the question is how best to estimate these liabilities when losses are not expected to emerge above deductible limits for many years.

Many questions still need to be addressed, for example:

• In the absence of long-term development histories under a deductible program, how can the actuary construct reasonable development factors? (Addressed in Section 3.)

• How can the actuary determine development patterns that reflect the diversity of both deductible size and mix? (Addressed in Section 3.)

• How should the actuary determine consistent development factors between limited and excess values? (Addressed in Section 3.)

• What is a reasonable approach for indexing deductible limits over time? (Addressed in Section 3.)

• How can the actuary estimate the liability associated with aggregate loss limits, if any? (Addressed in Section 5.1.)

• Is there a sound way to determine the proper asset value for associated service revenue? (Addressed in Section 5.2.)

1Similar in usage to a loss conversion factor in retro rating, loss multipliers are applied to deductible losses to capture expenses that vary with loss.
2 Development Approaches

2.1 Overview

The approaches discussed in this paper rely on my company's full coverage workers compensation claim experience. In effect, I create deductible/excess development patterns as needed. Further, the approach benefits from credible histories of full coverage losses, although the techniques used do not necessarily require a large volume of claim experience.

Once I establish development factors that reflect deductible amounts, I apply them at the account level and determine the overall aggregate reserve by summarizing estimated ultimates for each account. This is a reasonable approach if you view each account as belonging to a cohort of policies with similar limit characteristics. Determining the overall reserve allows me to address the issue of deductible mix by reflecting each account's limits.

In Section 4 I describe the possible use of a loss distribution model to enforce consistent results between deductible/excess development factors. Once the parameters of the distribution are set, it is possible to determine development factors for any deductible size. Such a model may provide an alternative approach for determining tail factors through the projection of the distribution parameters.

2.2 Loss Ratio

In the absence of credible development histories, a common approach for determining liabilities is to apply loss ratios to premiums arising from the exposures. As this element historically was required to price the product, loss ratios for the various accounts written should be readily available. As an alternative, the reserve actuary could use published loss ratios of workers compensation excess writers or reinsurers.

For immature years where data are sparse, applying loss ratios is the most practical approach to take. Given the long-tailed nature of this business, experience over deductible limits emerges slowly over time. Expected experience is readily converted to an accident year basis based upon a pro rata earnings of the policy year exposures.

The loss ratio approach requires a database of individual accounts and pricing elements. The database should include an estimate of the full coverage loss ratio. From a pricing standpoint, that estimate can
come from a variety of sources. One approach is to use company experience by state, reflecting the individual account's premium distribution. This experience possibly can be blended with industry experience. As with other pricing efforts, experience ought to be developed to ultimate, brought on level, and trended to the appropriate exposure period.

In addition to an estimate of the full coverage loss ratio, the database should include estimates of excess losses for both occurrence and aggregate limits. For the occurrence limit, several approaches are possible, including estimating excess ratios based upon company experience. A potentially more credible approach uses excess loss pure premium ratios underlying industry-based excess loss factors used in retro rating. Beside their availability by multiple limits, excess loss factors can easily be adjusted to a loss basis and to reflect hazard groups with differing severity potential. Using account-based excess ratios reflecting unique state and hazard group characteristics leads to reasonable estimates of per occurrence excess losses:

\[
\text{Per Occurrence Excess Losses} = P \times \text{ELR} \times \chi
\]

where

\[
P = \text{Premium;}
\]

\[
\text{ELR} = \text{Expected loss ratio; and}
\]

\[
\chi = \text{Per occurrence charge.}
\]

For the aggregate loss charge, I prefer a process similar to that used for determining insurance charges in a retro rating program. These charges rely on the National Council on Compensation Insurance's (NCCI) Table M. The process reflects the size of the account, deductible, state severity relativities, prospective rating period, and appropriate rating plan parameters:

\[
\text{Table M Aggregate Excess Losses} = P \times \text{ELR} \times (1 - \chi) \times \phi
\]

where \( \phi \) is the per aggregate charge.

Applying equations (1) and (2) to each account and then aggregating leads to an estimate of ultimate accident year losses. Table 1 shows

\(^2\)Actuaries are generally familiar with techniques used to determine loss ratios for pricing purposes, and a detailed description is beyond the scope of this paper.

\(^3\)I refer the interested reader to the Retrospective Rating Plan (1991) for further details.
a hypothetical case of how to apply both equations to determine the ultimate liabilities. Incurred but not reported (IBNR) amounts are determined by subtracting known losses from the ultimate estimate.

Again, the approach described in Table 1 is useful when no data are available or the data are too immature be useful. Loss ratio estimates can be consistently tied to pricing programs, at least at the outset. This approach also benefits from its reliance on a more credible pool of company and/or industry experience. On the negative side, the loss ratio approach has two shortcomings:

- It ignores emerging experience which may differ significantly from estimated ultimate losses. For this reason the loss ratio approach is not useful after several years of development; and
- It may not properly reflect account characteristics—development may emerge differently due to the exposures written.

2.3 Implied Development

There are many ways to incorporate emergence of losses in high deductible reserve estimates. Determining excess development implicitly is one possibility. The term implied development means an approach that works as follows:

- Develop full coverage losses to ultimate;
- Next develop deductible losses to ultimate by applying development factors reflecting various inflation indexed limits; and
- Determine ultimate excess losses by differencing the full coverage ultimate losses and the limited ultimate losses.

A variety of the usual development techniques can be applied to determine full coverage losses. These methods include paid and incurred techniques designed consistently with the company’s reserving procedures for full coverage workers compensation. Care should be exercised in determining a full coverage tail factor consistent with the limited loss tail factors. The actuary should avoid developing limited losses beyond unlimited losses or even losses for lower limits beyond those of higher limits.

When calculating development factors for the various deductibles, it is appropriate to index the limits for inflationary effects. Adjusting the deductible by indexing keeps the proportion of deductible/excess losses fairly constant about the limit from year to year.
### Table 1

**Countrywide Insurance Enterprise**  
**Account: Widget, Inc.**  
**Expected Deductible/Aggregate Loss Charges**

<table>
<thead>
<tr>
<th>State</th>
<th>Premium</th>
<th>ELR</th>
<th>Expected Loss</th>
<th>Excess Ratio</th>
<th>Deductible Loss Charge</th>
<th>Aggregate Ratio</th>
<th>Aggregate Loss Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas</td>
<td>9,084</td>
<td>0.567</td>
<td>5,151</td>
<td>0.062</td>
<td>319</td>
<td>0.02</td>
<td>97</td>
</tr>
<tr>
<td>Illinois</td>
<td>573,066</td>
<td>0.532</td>
<td>304,871</td>
<td>0.105</td>
<td>32,011</td>
<td>0.02</td>
<td>5,457</td>
</tr>
<tr>
<td>Iowa</td>
<td>373,072</td>
<td>0.588</td>
<td>219,366</td>
<td>0.096</td>
<td>21,059</td>
<td>0.02</td>
<td>3,966</td>
</tr>
<tr>
<td>Kansas</td>
<td>70,549</td>
<td>0.644</td>
<td>45,434</td>
<td>0.071</td>
<td>3,226</td>
<td>0.02</td>
<td>844</td>
</tr>
<tr>
<td>Minnesota</td>
<td>1,012,622</td>
<td>0.457</td>
<td>462,768</td>
<td>0.143</td>
<td>66,176</td>
<td>0.02</td>
<td>7,932</td>
</tr>
<tr>
<td>South Carolina</td>
<td>22,980</td>
<td>0.522</td>
<td>11,996</td>
<td>0.048</td>
<td>576</td>
<td>0.02</td>
<td>228</td>
</tr>
<tr>
<td>South Dakota</td>
<td>94,401</td>
<td>0.697</td>
<td>65,797</td>
<td>0.211</td>
<td>13,883</td>
<td>0.02</td>
<td>1,038</td>
</tr>
<tr>
<td>Total</td>
<td>2,155,774</td>
<td>0.517</td>
<td>1,115,383</td>
<td>0.123</td>
<td>137,250</td>
<td>0.02</td>
<td>19,562</td>
</tr>
</tbody>
</table>
For example, if inflationary forces drive claim costs 10 percent higher each year, one would expect the percentage of losses over a $100,000 deductible for one year equate to those of a $110,000 deductible in the next. Indexing of deductible limits allows for the possibility of combining differing experience years in the determination of development factors.

There is no set method for determining the indexing value. One approach is to determine the index by fitting a line to average severities over a long-term history. Another simpler approach is to use an index that reflects the movement in annual severity changes. The actuary needs to be cognizant that a constant deductible over time usually implies increasing excess losses.

An advantage of the implied development approach is that it provides an estimate of excess losses at early maturities when excess losses have not emerged. The development factors for limited losses are more stable than those determined for losses above the deductible. This approach also provides an important byproduct in the estimation of assets under the high deductible program. Estimating deductible losses helps determine the asset represented by revenue collected from the application of a loss multiplier to future losses. Despite these advantages, the implied development approach appears to misplace its focus by indirectly calculating excess losses, which can be problematic if one prefers to determine excess losses directly.

2.4 Direct Development

The direct development approach explicitly focuses on excess development, although it relies on development factors derived from the implied development approach. Given development factors for limited as well as full coverage losses, excess loss development factors can be calculated. Excess development is part of overall development, and the actuary should strive to determine excess factors in conjunction with limited development factors that balance to full coverage development. Reserve indications from implicit and explicit methods will not necessarily be the same, but the underlying loss development factors should be.

A variety of paid and incurred techniques are applicable here. I see several disadvantages to directly determining excess development factors and applying them to excess losses. These factors tend to be

---

4The excess losses are calculated indirectly by differencing ultimate unlimited losses with ultimate limited losses.
leveraged and extremely volatile, making them difficult to select. Additionally, if excess losses have not emerged at any particular stage of development, either the development factors do not exist or the indicated zero loss estimate is not particularly meaningful.

2.5 Credibility Weighting

There are significant drawbacks to the previous approaches (see Sections 2.2, 2.3, and 2.4) for determining excess liabilities. The credibility weighting approach offers more promise as it relies on credibility weighting indications based on experience with expected values, preferably based on pricing estimates. The actuary determines a suitable set of credibility weights then uses these weights to calculate the ultimate loss estimate \( \text{ULE}_t \), based on information at time \( t \):

\[
\text{ULE}_t = \text{OL}_t \times \text{LDF}_t \times Z_t + \text{EUL}_t \times (1 - Z_t)
\]

where

- \( \text{OL}_t \) = Observed loss at time \( t \);
- \( \text{LDF}_t \) = Age to ultimate development factor at time \( t \);
- \( Z_t \) = Credibility weight at time \( t \) and
- \( \text{EUL}_t \) = Expected ultimate loss at time \( t \).

The Bornhuetter-Ferguson (1972) technique offers one approach for determining credibility weights that are specified as reciprocals of loss development factors. For the Bornhuetter-Ferguson approach, substituting \( Z_t = 1/\text{LDF}_t \) into equation (3), yields:

\[
\text{ULE}_t = \text{OL}_t \times \text{ELR}_t \times \left( \frac{\text{LDF}_t - 1}{\text{LDF}_t} \right).
\]

Using the Bornhuetter-Ferguson approach allows the actuary to determine liabilities either directly or indirectly and can tie into pricing estimates for recent years where excess losses have yet to emerge. Also, it provides more stable estimates over time, rather than the volatility arising from erratic emergence or leveraged development factors. A credibility weighting approach such as this provides better estimators of ultimate liabilities as well. A disadvantage of the Bornhuetter-Ferguson approach is that a portion of \( \text{ULE}_t \), namely \( \text{ULE}_t \times (1 - Z_t) \), ignores observed losses. That drawback suggests finding alternative credibility weights that are more responsive to the actual experience as desired.
3 An Overview of Development Models

I will now deal more specifically with a number of questions raised in the introduction: How can the actuary determine development factors in the absence of a long-term history under the deductible program? How can the actuary determine development patterns that reflect the diversity of both deductible size and mix? What is a reasonable approach for indexing deductible limits over time? How should the process relate development for various limits consistently? As determining development factors for a high deductible program often is an exercise in partitioning development about the deductible limit, one question often is: Is it possible to develop consistent tail factors among the limits to which the company is exposed?

In the absence of long-term experience under the deductible program, I suggest using a company's history of full coverage workers compensation claims. It is also appropriate to apply an indexed limit to the claims to determine a series of accident year loss development histories by limit. I examine limits ranging from $50,000 to $1,000,000. I focus, however, on the more common deductible sizes in the neighborhood of $250,000. I use case losses including indemnity, medical, and any subject allocated claim expense. The histories run for 25 years but are not separated by account, injury, or state. This suggests creating alternative development patterns that reflect these factors. Table 2 shows age to age development factors by indexed limit.

Table 2
Workers Compensation—High Deductibles
Limited Loss and ALAE Age to Age Development
Factors by Indexed Limit (Middle Six of Last Eight)

<table>
<thead>
<tr>
<th>Limit</th>
<th>12:24</th>
<th>24:36</th>
<th>36:48</th>
<th>48:60</th>
<th>60:72</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,000</td>
<td>1.5031</td>
<td>1.0418</td>
<td>1.0038</td>
<td>1.0025</td>
<td>1.0020</td>
</tr>
<tr>
<td>$100,000</td>
<td>1.6225</td>
<td>1.0727</td>
<td>1.0151</td>
<td>1.0063</td>
<td>1.0080</td>
</tr>
<tr>
<td>$250,000</td>
<td>1.6791</td>
<td>1.1300</td>
<td>1.0451</td>
<td>1.0207</td>
<td>1.0060</td>
</tr>
<tr>
<td>$500,000</td>
<td>1.6827</td>
<td>1.1393</td>
<td>1.0684</td>
<td>1.0322</td>
<td>1.0170</td>
</tr>
<tr>
<td>$750,000</td>
<td>1.6816</td>
<td>1.1408</td>
<td>1.0720</td>
<td>1.0359</td>
<td>1.0214</td>
</tr>
<tr>
<td>$1,000,000</td>
<td>1.6811</td>
<td>1.1411</td>
<td>1.0728</td>
<td>1.0371</td>
<td>1.0229</td>
</tr>
<tr>
<td>Unlimited</td>
<td>1.6876</td>
<td>1.1430</td>
<td>1.0749</td>
<td>1.0391</td>
<td>1.0196</td>
</tr>
</tbody>
</table>
In order to determine those development factors, I combine several years of experience based upon indexed limits. For example, for the most recent year limits are used as stated. But for the first prior year I adjust limits downward by an indexing factor of 1.095. For the current year I assume a limit of $250,000 was the equivalent of a limit of $228,311 for the first prior year. Each limit is adjusted by the same index to generate the desired development factors. Figure 1 shows the exponential trend line fit through known data points determining the long-term indexing factor of 1.095. Also depicted is the indexed $250,000 loss limit.

I recommend separating claim count development from severity development when estimating high deductible liabilities. In this paper I focus on the counts for full coverage losses rather than the emergence of claims over a specific deductible limit. It is easier to recognize development in this fashion, as there is generally little true claim count IBNR after three years. This is true for larger claims, as they will be reported early (like other claims), but their severity will not be known for some time.

Table 3
Workers Compensation—Age to Age Development Factors

<table>
<thead>
<tr>
<th>Full Coverage Claim Count</th>
<th>Months:Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident Year</td>
<td>12:24 24:36 36:48 48:60</td>
</tr>
<tr>
<td>1988</td>
<td>- - - 0.9999</td>
</tr>
<tr>
<td>1989</td>
<td>- - 0.9999 0.9994</td>
</tr>
<tr>
<td>1990</td>
<td>- 1.0026 0.9999 1.0001</td>
</tr>
<tr>
<td>1991</td>
<td>1.1111 1.0022 1.0002 -</td>
</tr>
<tr>
<td>1992</td>
<td>1.1305 1.0017 - -</td>
</tr>
<tr>
<td>1993</td>
<td>1.1283 - - -</td>
</tr>
<tr>
<td>Last Three</td>
<td>1.1233 1.0022 1.0000 0.9998</td>
</tr>
<tr>
<td>Selected</td>
<td>1.1250 1.0025 1.0000 1.0000</td>
</tr>
<tr>
<td>Age to Ultimate</td>
<td>1.1278 1.0025 1.0000 1.0000</td>
</tr>
</tbody>
</table>

To develop limited losses to ultimate, I use a three parameter version of the inverse power curve recommended by Sherman (1984) to

\[\text{The selected indexing factor of 1.095 is based upon a long-term severity history. There may be, however, better approaches such as varying the indexing factor by year or adjusting for the distorting effects of larger claims.}\]
model the development arising in the tail. The curve can be written as a function of $t$ as follows:

$$y(t) = 1 + \frac{a}{(t + c)^b} \quad t > 0,$$

(5)

where $a$, $b$, and $c$ are constants. My concern is to determine consistent tail factors by limit. Starting with the unlimited loss development and fitting an inverse power curve to known age to age factors allows me to project ultimate unlimited losses. As the inverse power curve is defined for $t > 0$, a time to end the projection must be selected. I use a method that relies on extended development triangles. (The method is similar to the method used for determining our full coverage tail factor.) It turns out that the projected age to age development factors can be stopped at 40 years. Compounding the age to age factors from the fitted curve leads to the desired completion or tail factors.

Once the ultimate age is determined, I use a minimum chi-square (between observed and expected values) technique to fit inverse power curves to the age to age factors for the various deductible limits and extend to the common maturity. Although this approach generates
uniformly decreasing tail factors, it is not clear whether the bias in extending all curves to a common maturity is significant. (At lower limits, development likely ceases before 40 years.) Figure 2 depicts the age to age model determined for the unlimited loss development. Figure 3 shows the pattern of age to ultimate limited loss development factors resulting from the inverse power curve model.

Another issue is the relationship between loss development factors and limited severity relativities. In some of my earlier efforts I attempted to develop losses by limit without regard to how they might relate to one another. This led to inconsistencies in development factors where completion factors for smaller deductibles, for example, sometimes exceeded factors for larger deductibles. I found that any attempts to determine deductible development factors need to address the relationship between the full coverage loss development and limited severity relativities.

---

6Limited severity relativities are defined simply as the ratio of the limited to unlimited severity.
The following formulas show limited development factors ($LDF^L_t$) and excess development factors ($XSLDF^L_t$) at time $t$ as functions of the unlimited loss development and limited severity relativities:

$$LDF^L_t = \frac{C}{C_t} \times \frac{S}{S_t} \times \frac{R^L_t}{R^L}$$ \hspace{1cm} (6)$$

$$XSLDF^L_t = \frac{C}{C_t} \times \frac{S}{S_t} \times \frac{1 - R^L_t}{1 - R^L}$$ \hspace{1cm} (7)$$

where $t$ is the age and

$L = \text{Deductible limit};$

$C = \text{Ultimate claim count};$

$C_t = \text{Total claim count at time } t;$

$S = \text{Ultimate full coverage severity};$

$S_t = \text{Full coverage severity at time } t;$
\[ R \quad \text{Ultimate limited severity relativity; and} \]
\[ R_t \quad \text{Limited severity relativity at time } t. \]

The motivation for equations (6) and (7) results from the desire to partition total loss development in a consistent fashion between limited and excess development. Note that

\[
LDF_t^L = R_t^L \times LDF_t^L + (1 - R_t^L) \times XSLDF_t^L
\]

\[
= R_t^L \times \frac{C}{C_t} \times \frac{S}{R_t^L} + (1 - R_t^L) \times \frac{C}{C_t} \times \frac{S}{S_t} \times \frac{1 - R_t^L}{1 - R_t^L}
\]

\[
= \frac{C}{C_t} \times \frac{S}{S_t}
\]

as is expected.

Figure 4 shows how the historical limited severity relativities ought to relate to each other and how they change over time. The relativity curves cluster near unity initially and progressively decrease over time for smaller and smaller deductibles without crossing over one another.

**Figure 4**

**Workers Compensation**

**High Deductibles Limited Severity Relativities**

Table 4 shows age to age development about a $250,000 deductible limit. Table 5 shows relativities and their changes for the selected deductible limit. Note how the change in limited loss development relates
to the unlimited loss development. Actual case loss development does not always conform to expectations, as the limited loss development factor sometimes exceeds the unlimited, thus

$$LDF_t^L = LDF_t \times \Delta R_t^L$$  \hspace{1cm} (9)$$

where $\Delta f_t^L = f_t^L / f_t$ for any function $f$. For example, for accident year 1993 moving from 12 to 24 months, a limited factor of 1.6229 is observed. This is equivalent to the unlimited loss development factor of 1.6044 compounded with the change in severity relativities for the same time period of 1.0116.

Note also the relationship of the excess development to the unlimited loss development for the same year:

$$XSLDF_t^L = LDF_t \times \Delta(1 - R_t^L).$$  \hspace{1cm} (10)$$

The accident year 1993 excess development factor of 1.1684 is equivalent to the unlimited development factor times the ratio of the complements of the severity relativities moving from 12 to 24 months, i.e., $1.1684 = 1.6044 \times (1 - 0.9704)/(1 - 0.9593)$. The weighted average of the limited and excess development factors using the relativity as a weight gives the unlimited loss development factor in equation (7). Also $1.6044 = 0.9704 \times 1.6229 + (1 - 0.9704) \times 1.1684$ for accident year 1993.

## 4 Distributional Models

Statistical distributions are ideally suited for modeling loss development factors as they can tie the relativities to the severities and consequently provide consistent loss development factors. They model the development process by determining parameters of a distribution that vary over time.\(^7\)

Once the distribution and its parameters are specified, it is possible to calculate the desired limited/excess severities. Comparing those severities over time leads to the needed development factors. Care has to be exercised to recognize claim count development at earlier maturities. Also, distributional models allow for interpolation among limits and years as needed.

---

\(^7\)I use experience in the modeling process for known points in time. I rely upon techniques described previously to determine the projected ultimate losses for the final point in time.
### Table 4
Workers Compensation—High Deductibles
Age to Age Loss and ALAE Development Factors

<table>
<thead>
<tr>
<th>Accident Year</th>
<th>Month:Month</th>
<th>Panel A: Unlimited</th>
<th>Panel B: $250,000 Deductible</th>
<th>Panel C: Excess of $250,000 Deductible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12:24</td>
<td>24:36</td>
<td>36:48</td>
<td>48:60</td>
</tr>
<tr>
<td>1989</td>
<td>1.7063</td>
<td>1.1756</td>
<td>1.0929</td>
<td>1.0359</td>
</tr>
<tr>
<td>1990</td>
<td>1.8219</td>
<td>1.1574</td>
<td>1.0744</td>
<td>1.0387</td>
</tr>
<tr>
<td>1991</td>
<td>1.7724</td>
<td>1.1506</td>
<td>1.0737</td>
<td>-</td>
</tr>
<tr>
<td>1992</td>
<td>1.6912</td>
<td>1.1398</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1993</td>
<td>1.6044</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>1.7192</td>
<td>1.1559</td>
<td>1.0803</td>
<td>1.0373</td>
</tr>
</tbody>
</table>

| 1989          | 1.7077      | 1.1598             | 1.0657                        | 1.0221                                | 1.0120                                |
| 1990          | 1.7755      | 1.1509             | 1.0550                        | 1.0247                                | -                                     |
| 1991          | 1.7734      | 1.1461             | 1.0643                        | -                                     | -                                     |
| 1992          | 1.6750      | 1.1363             | -                             | -                                     | -                                     |
| 1993          | 1.6229      | -                  | -                             | -                                     | -                                     |
| Average       | 1.7109      | 1.1483             | 1.0617                        | 1.0234                                | 1.0120                                |

| 1989          | 1.6646      | 1.6582             | 1.6742                        | 1.1927                                | 1.2011                                |
| 1990          | 4.4890      | 1.3049             | 1.3151                        | 1.2411                                | -                                     |
| 1991          | 1.7373      | 1.3115             | 1.3675                        | -                                     | -                                     |
| 1992          | 2.2474      | 1.2291             | -                             | -                                     | -                                     |
| 1993          | 1.1684      | -                  | -                             | -                                     | -                                     |
| Average       | 2.2613      | 1.3759             | 1.4523                        | 1.2169                                | 1.2011                                |
Table 5

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Month</th>
<th>Month</th>
<th>Month</th>
<th>Month</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>12</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td>60</td>
<td>72</td>
</tr>
<tr>
<td>1990</td>
<td>0.9683</td>
<td>0.9623</td>
<td>0.9594</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1991</td>
<td>0.9704</td>
<td>0.9593</td>
<td>0.9553</td>
<td>0.9315</td>
<td>0.9191</td>
<td>0.9053</td>
</tr>
<tr>
<td>1992</td>
<td>0.9717</td>
<td>0.9675</td>
<td>0.9532</td>
<td>0.9333</td>
<td>0.9191</td>
<td>0.9053</td>
</tr>
<tr>
<td>1993</td>
<td>0.9593</td>
<td>0.9723</td>
<td>0.9675</td>
<td>0.9424</td>
<td>0.9227</td>
<td>0.9053</td>
</tr>
<tr>
<td>Average</td>
<td>0.9707</td>
<td>0.9663</td>
<td>0.9590</td>
<td>0.9424</td>
<td>0.9209</td>
<td>0.9053</td>
</tr>
</tbody>
</table>

Panel A: Relativities

Year | Month:Month
---|---
1989 | 12:24
1990 | 24:36
1991 | 36:48
1992 | 48:60
1993 | 60:72

Panel B: Changes in Relativities

Year | Month:Month
---|---
1989 | 12:24
1990 | 24:36
1991 | 36:48
1992 | 48:60
1993 | 60:72

I use a Weibull distribution to specify the workers compensation claim loss distribution. The Weibull distribution is commonly used for workers compensation claims because it gives a reasonable depiction of the loss distributions. Some of the properties of the Weibull distribution are given in Hogg and Klugman (1984, Appendix, page 231). The cumulative distribution function (cdf), probability distribution function (pdf), moments and the truncated mean are shown below:

\[ F(x) = 1 - e^{-(x/\beta)\alpha}, \quad x > 0 \]  
\[ f(x) = \frac{\alpha x^{\alpha-1} e^{-(x/\beta)\alpha}}{\beta^\alpha}, \quad x > 0 \]  
\[ E[X^n] = \beta^n \Gamma(1 + \frac{n}{\alpha}) \text{ for } n = 1, 2, \ldots \]
\[ \text{LEV}[X; d] = \int_{0}^{d} x f(x) \, dx + d \times [1 - F(x)] \]
\[ d \varepsilon^{-\left(\frac{d}{\beta}\right)^\alpha} = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) P\left(1 + \frac{1}{\alpha}; \left(\frac{d}{\beta}\right)^\alpha\right) + d e^{-\left(\frac{d}{\beta}\right)^\alpha}, \quad d > 0\]

where \( \alpha > 0 \) is the shape parameter, and \( \beta > 0 \) is the scale parameter. In addition, for \( a > 0 \), \( \Gamma(a) \) is the gamma function, and \( P(a, x) \) is the incomplete gamma function, i.e.,

\[
\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt = (a - 1)\Gamma(a - 1)
\]
\[
P(a, x) = \int_0^x t^{a-1} e^{-t} dt.
\]

For accurate approximations to \( \Gamma(a) \) and \( P(a, x) \), see Abramowitz and Stegun (1965, Chapter 6).

The most difficult aspect of working with distributional models is estimating the unknown parameters involved. There are various statistical approaches that can be used, including the method of moments and the maximum likelihood estimation. I use an alternative approach called the minimum chi-square, which is based on the minimization of the sum of the squared deviations between actual and expected severity relativities around the $250,000 deductible size.

\[
\chi^2 = \min_{\alpha, \beta} \left[ \sum_i \frac{(\text{Actual}_i - \text{Expected}_i)^2}{\text{Expected}_i} \right].
\]

The estimates of \( \alpha \) and \( \beta \) are the parameter values that actually minimize chi-square (\( \chi^2 \)). I use a solver routine incorporated in Microsoft Excel's spreadsheet application, which allows me to constrain the optimization routine in such a fashion that the parameters generated produced the actual unlimited severity at the specified maturity.

Table 6 shows an example of results used to determine age to ultimate loss development factors by limit from 48 months to ultimate. In the table the limited and excess severities sum to the unlimited severity, as I base all severities upon total claim counts. I select 48 months to focus attention on changes in severity rather than changes in total claim counts assuming no IBNR count development after 36 months.

The following formulation shows how expected development at time \( t \), \( ED_t \) can be partitioned about the deductible limit:
Table 6
Workers Compensation High Deductibles Actual Versus Fitted Limited/Excess Development Factors (at 48 Months) Using a Weibull Loss Distribution

<table>
<thead>
<tr>
<th>Limit</th>
<th>Unlimited</th>
<th>$1,000,000</th>
<th>$750,000</th>
<th>$500,000</th>
<th>$250,000</th>
<th>$100,000</th>
<th>$50,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Ultimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Severity</td>
<td>6,846.4</td>
<td>6,159.2</td>
<td>5,980.4</td>
<td>5,714.4</td>
<td>5,094.8</td>
<td>3,939.6</td>
<td>3,036.5</td>
</tr>
<tr>
<td>Relativity</td>
<td>1.0000</td>
<td>0.8996</td>
<td>0.8735</td>
<td>0.8347</td>
<td>0.7442</td>
<td>0.5754</td>
<td>0.4435</td>
</tr>
<tr>
<td>Excess Severity</td>
<td>0.0</td>
<td>687.2</td>
<td>866.0</td>
<td>1,132.0</td>
<td>1,751.6</td>
<td>2,906.8</td>
<td>3,809.9</td>
</tr>
<tr>
<td>Fitted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Severity</td>
<td>6,846.4</td>
<td>6,295.2</td>
<td>6,106.5</td>
<td>5,778.7</td>
<td>5,064.4</td>
<td>3,926.7</td>
<td>3,043.8</td>
</tr>
<tr>
<td>Relativity</td>
<td>1.0000</td>
<td>0.9195</td>
<td>0.8919</td>
<td>0.8440</td>
<td>0.7397</td>
<td>0.5735</td>
<td>0.4446</td>
</tr>
<tr>
<td>Excess Severity</td>
<td>0.0</td>
<td>551.2</td>
<td>739.9</td>
<td>1,067.7</td>
<td>1,782.0</td>
<td>2,919.7</td>
<td>3,802.6</td>
</tr>
<tr>
<td>Weibull Parameters: Scale = 180.0, Shape = 0.2326, Mean = 6,846.4, Coefficient of Variation = 10.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: 48 Months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Severity</td>
<td>5,530.2</td>
<td>5,346.6</td>
<td>5,288.5</td>
<td>5,182.3</td>
<td>4,824.0</td>
<td>3,807.5</td>
<td>2,937.1</td>
</tr>
<tr>
<td>Relativity</td>
<td>1.0000</td>
<td>0.9668</td>
<td>0.9563</td>
<td>0.9371</td>
<td>0.8723</td>
<td>0.6885</td>
<td>0.5311</td>
</tr>
<tr>
<td>Limited LDF</td>
<td>1.2380</td>
<td>1.1520</td>
<td>1.1308</td>
<td>1.1027</td>
<td>1.0561</td>
<td>1.0347</td>
<td>1.0338</td>
</tr>
<tr>
<td>Excess Severity</td>
<td>0.0</td>
<td>183.6</td>
<td>241.7</td>
<td>347.9</td>
<td>706.2</td>
<td>1,722.7</td>
<td>2,593.1</td>
</tr>
<tr>
<td>Excess LDF</td>
<td>-</td>
<td>3.7429</td>
<td>3.5830</td>
<td>3.2538</td>
<td>2.4803</td>
<td>1.6874</td>
<td>1.4692</td>
</tr>
<tr>
<td>Fitted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Severity</td>
<td>5,530.2</td>
<td>5,380.5</td>
<td>5,301.4</td>
<td>5,142.5</td>
<td>4,722.4</td>
<td>3,894.0</td>
<td>3,144.1</td>
</tr>
<tr>
<td>Relativity</td>
<td>1.0000</td>
<td>0.9729</td>
<td>0.9586</td>
<td>0.9299</td>
<td>0.8539</td>
<td>0.7041</td>
<td>0.5685</td>
</tr>
<tr>
<td>Limited LDF</td>
<td>1.2380</td>
<td>1.1700</td>
<td>1.1519</td>
<td>1.1237</td>
<td>1.0724</td>
<td>1.0084</td>
<td>0.9681</td>
</tr>
<tr>
<td>Excess Severity</td>
<td>0.0</td>
<td>149.7</td>
<td>228.8</td>
<td>387.7</td>
<td>807.8</td>
<td>1,636.2</td>
<td>2,386.1</td>
</tr>
<tr>
<td>Excess LDF</td>
<td>-</td>
<td>3.6820</td>
<td>3.2338</td>
<td>2.7539</td>
<td>2.2060</td>
<td>1.7844</td>
<td>1.5936</td>
</tr>
<tr>
<td>Weibull Parameters: Scale = 305.7, Shape = 0.2625, Mean = 5,530.2, Coefficient of Variation = 7.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The first part of equation (11) represents the portion of expected development below the deductible limit (i.e., deductible development), while the second part of equation (11) represents the portion of expected development above the deductible limit (i.e., excess development). Figure 5 shows partitioned development above (excess) and below (deductible) a $250,000 deductible limit based upon the Weibull loss distribution model. Excess development represents the majority of development with increasing age.

Figure 5
Workers Compensation—High Deductibles
Partitioned Development Above & Below $250,000 Deductible Limit
5 Other Elements

Several other elements associated with high deductible plans call for further discussion: aggregate limits, service revenue, and allocated claim expense. Determining sound estimates for these items is complex. In the following discussion I recommend using the compound Poisson model of collective risk to estimate losses excess of aggregate limits. I also suggest an alternative procedure using the NCCI Table M, if collective risk modeling is impractical. The asset for service revenue, although not as difficult to determine, depends upon prior estimates of losses for deductible/aggregate limits. Treating allocated claim expense in a similar fashion to loss simplifies the estimation process for that liability, but separating the two pieces is problematic.

5.1 Aggregate Limits

Some risks, beside choosing to limit their per occurrence losses, desire to limit all losses under a high deductible program. Insurers satisfy that need by providing aggregate loss limits. These limits are conceptually similar to maximum premium limitations used in retrospective rating plans.

Determining loss development factors for losses excess of aggregate limits is more complicated than for per occurrence limitations. The obligations from these aggregate limits are generally less significant than for per occurrence limits. Beside the additional complexity, the data needed to determine development factors for these limits are generally sparse and not likely to be credible. Outside of attempting to gather data for development factors, I suggest using collective risk modeling techniques to determine the needed loss development factors. Specifically, I use the Heckman and Meyers (1983) collective risk model with a Poisson claim frequency distribution and a Weibull claim severity distribution to determine development factors for losses excess of aggregate limits. Table 7 shows selected development factors using the Weibull loss distribution.

The sampling of development factors shows that development for losses in excess of aggregate limits decreases more rapidly over time with smaller deductibles than larger ones. This is not unexpected, as most of the later development occurs in the layers of loss above the

8 Although I do not incorporate any parameter risk in determining the development factors, the model does allow for that possibility. Interested readers should see a discussion by Meyers and Schenker (1983) describing how to incorporate parameter risk into the collective risk model.
deductible limits which is not covered by the aggregate. Also, not un-expectedly, development is more leveraged for larger aggregate limits. There is one additional point the reader should note in reviewing Table 7. Although I show hypothetical results for risks of $1 million and $2.5 million in expected loss size, the limited expectations are considerably smaller.

Given the volatility of losses excess of aggregate limits, I recommend using a Bornhuetter-Ferguson method to smooth indications of ultimate liability. The example in Table 8 uses expected aggregate loss charges as well as expected development factors based on the collective risk modeling approach. The final indication adds known losses excess of aggregate limits and IBNR based on the modeled development patterns.

An alternative approach for determining IBNR estimates for aggregate excess of loss coverage merits consideration. The procedure uses the NCCI methodology (1991) for determining insurance charges in retrospective rating plans. It is a more practical approach than collective risk modeling, but its accuracy hinges on determining the proper insurance charge table.

The IBNR is determined by subtracting insurance charges at different maturities. The process used to determine the ultimate insurance charge is the same as that used for pricing purposes. The key to the NCCI procedure is the adjustment of expected losses reflecting loss limits. This adjustment increases expected losses used in determining the appropriate insurance charge table using the following formula:

\[
\text{Adjustment Factor} = \frac{1 + 0.8X}{1 - X}.
\]

The reason for increasing expected losses for the use of a per occurrence limit is to use a less dispersed loss ratio distribution and, consequently, a smaller insurance charge. Although this adjustment for a loss limit moves the selection of an insurance charge table in the right direction, we are not sure if the move has been made in an appropriate manner. Additionally, the procedure generates smaller insurance charges by using limited losses in the entry ratio calculation.

In order to calculate the insurance charge at earlier maturities I suggest determining the per occurrence charge used in the NCCI procedure by relating undeveloped, limited losses to ultimate, unlimited losses. For example, using the fitted results depicted in Table 6 for a $250,000 deductible leads to a per occurrence charge of 31 percent \((1 - 4722.4/6846.4)\) at 48 months.
Table 7
Workers Compensation High Deductibles
Development Factors for Losses Excess of Aggregate Limits
(Collective Risk Model Using Weibull Loss Distribution)

<table>
<thead>
<tr>
<th>Aggregate Limit</th>
<th>Deductible</th>
<th>12 Months</th>
<th></th>
<th>48 Months</th>
<th></th>
<th>Ultimate Excess Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Excess Loss</td>
<td>LDF</td>
<td>Excess Loss</td>
<td>LDF</td>
<td></td>
</tr>
<tr>
<td>Panel A: Expected Unlimited Losses of $1,000,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$500,000</td>
<td>$100,000</td>
<td>9,253.6</td>
<td>13.024</td>
<td>114,646.0</td>
<td>1.051</td>
<td>120,523.3</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>22,882.5</td>
<td>12.007</td>
<td>228,070.7</td>
<td>1.205</td>
<td>274,761.6</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>28,653.6</td>
<td>13.255</td>
<td>289,389.2</td>
<td>1.312</td>
<td>379,794.3</td>
</tr>
<tr>
<td>$750,000</td>
<td>$100,000</td>
<td>155.1</td>
<td>136.451</td>
<td>84,475.1</td>
<td>1.394</td>
<td>117,788.5</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>1,844.9</td>
<td>49.763</td>
<td>138,526.3</td>
<td>1.529</td>
<td>211,851.8</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>4,257.2</td>
<td>47.63</td>
<td>213,725.7</td>
<td>1.835</td>
<td>306,854.6</td>
</tr>
<tr>
<td>$1,000,000</td>
<td>$100,000</td>
<td>0.8</td>
<td>2,242.750</td>
<td>1,274.7</td>
<td>1.408</td>
<td>1,794.2</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>94.5</td>
<td>418.531</td>
<td>23,343.1</td>
<td>1.694</td>
<td>39,551.2</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>494.5</td>
<td>213.275</td>
<td>57,471.2</td>
<td>1.835</td>
<td>105,464.6</td>
</tr>
<tr>
<td>Panel B: Expected Unlimited Losses of $2,500,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1,000,000</td>
<td>$100,000</td>
<td>39,703.2</td>
<td>11.761</td>
<td>456,498.9</td>
<td>1.023</td>
<td>466,934.1</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>81,084.7</td>
<td>10.876</td>
<td>759,354.4</td>
<td>1.161</td>
<td>881,844.0</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>95,069.6</td>
<td>12.021</td>
<td>912,976.1</td>
<td>1.252</td>
<td>1,142,866.6</td>
</tr>
<tr>
<td>1,250,000</td>
<td>$100,000</td>
<td>3,829.0</td>
<td>64.779</td>
<td>236,271.2</td>
<td>1.050</td>
<td>248,037.5</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>17,740.7</td>
<td>36.191</td>
<td>522,364.3</td>
<td>1.229</td>
<td>642,046.5</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>26,520.1</td>
<td>33.986</td>
<td>674,759.3</td>
<td>1.336</td>
<td>901,315.4</td>
</tr>
<tr>
<td>1,500,000</td>
<td>$100,000</td>
<td>173.5</td>
<td>564.077</td>
<td>87,988.1</td>
<td>1.112</td>
<td>97,867.3</td>
</tr>
<tr>
<td></td>
<td>$250,000</td>
<td>2,693.1</td>
<td>158.522</td>
<td>318,464.5</td>
<td>1.341</td>
<td>426,916.3</td>
</tr>
<tr>
<td></td>
<td>$500,000</td>
<td>6,001.8</td>
<td>112.833</td>
<td>463,359.8</td>
<td>1.461</td>
<td>677,200.3</td>
</tr>
</tbody>
</table>
Table 8
Countrywide Insurance Enterprise
Workers Compensation—High Deductibles
Estimated Ultimate Aggregate Excess of Loss
(Utilizing Bornhuetter-Ferguson Methodology)

<table>
<thead>
<tr>
<th>Panel A: EDL = $1,000,000; AGL = $750,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 100,000 581,252 - 21,164 1.175 3,152</td>
</tr>
<tr>
<td>B 250,000 703,027 - 117,789 1.394 33,292</td>
</tr>
<tr>
<td>C 500,000 764,493 14,493 211,852 1.529 87,789</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: EDL = $2,500,000; AGL = $1,250,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 100,000 1,453,169 203,169 248,038 1.050 214,980</td>
</tr>
<tr>
<td>Y 250,000 1,757,616 507,616 642,047 1.229 627,248</td>
</tr>
<tr>
<td>Z 500,000 1,911,285 661,285 901,315 1.336 887,963</td>
</tr>
</tbody>
</table>

EUL = Expected Unlimited Loss; AGL = Aggregate Limit; EXAG = Excess of Aggregate
ACCT = Account; DEDUCT = Deductible

In addition to reflecting the impact of the limit, this approach also captures the effects of development. Again, the issue is whether the adjustments for the limit and for development are appropriate.

Table 9 compares IBNR estimates determined using the NCCI Table M with estimates from the collective risk modeling approach depicted in Table 8. Table 10 contains further details for the IBNR estimates from the NCCI Table M procedure.

5.2 Service Revenue

A significant element that ought to be reflected on the asset side of the balance sheet is the revenue associated with servicing claims under a high deductible program. Service revenue is generated in an analogous fashion to the loss conversion factor in a retro rating plan. Generally, a factor is applied to deductible losses, limited by any applicable aggregate, to cover expenses that vary with these losses.

In practice, however, other elements are captured by the loss multiplier, reflecting the desire of the individual accounts to fund the cost of the program as losses emerge. The service revenue often is collected
Table 9
A Comparison of Aggregate Excess of Loss IBNR Estimates at 48 Months
Collective Risk Model vs. NCCI Table M

<table>
<thead>
<tr>
<th>Account</th>
<th>Deductible</th>
<th>Collective Risk Model</th>
<th>NCCI Table M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: EUL = $1,000,000; AGL = $750,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>100,000</td>
<td>3,152</td>
<td>1,809</td>
</tr>
<tr>
<td>B</td>
<td>250,000</td>
<td>33,292</td>
<td>38,500</td>
</tr>
<tr>
<td>C</td>
<td>500,000</td>
<td>73,296</td>
<td>68,811</td>
</tr>
<tr>
<td>Panel B: EUL = $2,500,000; AGL = $1,250,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>100,000</td>
<td>11,811</td>
<td>9,959</td>
</tr>
<tr>
<td>Y</td>
<td>250,000</td>
<td>119,633</td>
<td>103,000</td>
</tr>
<tr>
<td>Z</td>
<td>500,000</td>
<td>226,678</td>
<td>222,168</td>
</tr>
</tbody>
</table>

EUL = Expected Ultimate Loss; AGL = Aggregate Limit

as losses are paid, but it also may be gathered as a function of case-incurred losses.

I propose determining the asset in the following fashion:

- Determine ultimate deductible losses at the account level;
- Subtract ultimate losses excess of aggregate limits from ultimate deductible losses;
- Apply the selected loss multiplier to the difference determined in the second step to determine ultimate recoverables; and
- Determine the total asset by subtracting any known recoveries from the estimated ultimate recoverables and aggregate results for all accounts.

Table 11 shows an example of how in practice the asset for the service revenue may be determined.
Table 10
Determination of IBNR—Aggregate Excess of $1,250,000

<table>
<thead>
<tr>
<th>Item</th>
<th>At 48 Months</th>
<th>Ultimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Severity Deductible = L = $250,000</td>
<td>4722.4</td>
<td>5064.4</td>
</tr>
<tr>
<td>(b) Frequency</td>
<td>365.2</td>
<td>365.2</td>
</tr>
<tr>
<td>(c) Limited Loss: (a) × (b)</td>
<td>1,724,620.5</td>
<td>1,849,518.9</td>
</tr>
<tr>
<td>(d) Entry Ratio: Aggregate</td>
<td>0.72</td>
<td>0.68</td>
</tr>
<tr>
<td>(e) Loss Excess of Deductible: 1 - LEV(X;L) / E[X]</td>
<td>0.310</td>
<td>0.260</td>
</tr>
<tr>
<td>(f) Adjustment for Limit: (1 + 0.8 × (e)) / (1 - (e))</td>
<td>1.810</td>
<td>1.633</td>
</tr>
<tr>
<td>(g) Adjusted Limited Loss: EU × (f)</td>
<td>4,525,000</td>
<td>4,082,500</td>
</tr>
<tr>
<td>(h) 1994 Expected Loss Group</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>(i) Insurance Charge Ratio</td>
<td>0.336</td>
<td>0.369</td>
</tr>
<tr>
<td>(j) Insurance Charge Amount: (c) × (i)</td>
<td>579,472</td>
<td>682,472</td>
</tr>
<tr>
<td>(k) IBNR - 682,472 - 579,472</td>
<td>103,000</td>
<td></td>
</tr>
</tbody>
</table>

Risk Characteristics: Expected Unlimited Loss = $2,500,000; Severity = 6846.4; and Frequency = 365.2
<table>
<thead>
<tr>
<th>Account</th>
<th>Deductible</th>
<th>Excess of Aggregate</th>
<th>Net of Aggregate</th>
<th>Multiplier Revenue</th>
<th>Known Recoveries</th>
<th>Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Expected Unlimited Loss - $2,500,000; Aggregate Limit - $1,250,000; Loss Multiplier - 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>1,465,376</td>
<td>214,980</td>
<td>1,250,396</td>
<td>125,040</td>
<td>96,960</td>
<td>28,080</td>
</tr>
<tr>
<td>Y</td>
<td>1,884,867</td>
<td>627,248</td>
<td>1,257,619</td>
<td>125,762</td>
<td>102,712</td>
<td>23,050</td>
</tr>
<tr>
<td>Z</td>
<td>2,147,711</td>
<td>887,963</td>
<td>1,259,748</td>
<td>125,975</td>
<td>106,912</td>
<td>19,063</td>
</tr>
<tr>
<td>Total</td>
<td>5,497,954</td>
<td>1,730,191</td>
<td>3,767,763</td>
<td>376,777</td>
<td>306,584</td>
<td>70,193</td>
</tr>
</tbody>
</table>
5.3 Allocated Claim Expense

There are two principal means of handling allocated claim expense under a high deductible program. Either the account manages this expense, or the expense is treated as loss and subjected to applicable limits. In the first instance development patterns reflecting loss only are appropriate for determining liabilities, while a combination of loss and expense is appropriate for the second case.

For this discussion I determine development factors by combining loss and expense components assuming expenses developed similar to losses. I apply the resulting development factors to experience arising from both types of plans, as most of the accounts we write choose to subject allocated claim expense to the deductible. Although different development patterns are likely for loss and expense versus loss only, the actuary needs to decide based upon the mix of plans whether using both development patterns is worth the effort.

A remaining issue is how best to split loss and allocated claim expense for financial reporting purposes. Although splitting them proportionately based upon their full coverage counterparts is expeditious, other more actuarially sound approaches, even if available, may not be cost justifiable.

6 Conclusion

This discussion suggests some possible approaches for estimating liabilities under a high deductible program. As with many actuarial procedures, much work and improvement still are needed. I hope my suggestions provoke further discussion about how to better estimate these liabilities.

Although the reader may have his or her own ideas on how to improve upon my suggestions, I think several of the following suggestions warrant further consideration:

- Obtain longer histories of experience under the program better reflecting risk characteristics;
- Derive (select) parameters (distributions) that provide better fits to the actual data;
- Determine better tail factors and/or parameters of the utilized loss distribution; and
- Develop more advanced approaches to index loss limits.
These are known issues for actuaries, who long have confronted either limited or excess loss development. More comprehensive data in a workers compensation program allows the application of more sophisticated loss distributional approaches which affords greater consistency to all of the pieces involved. To that end I hope this paper provides a few steps toward developing sounder actuarial techniques for analyzing workers compensation high deductible loss development.

References


Third Party Administrator (TPA) Service Pricing and Incentive Contracts

Hou-Wen Jeng*

Abstract†

This paper addresses a few of the most important pricing issues faced by a third party administrator (TPA) whose main responsibility is claims handling for self-insured employers and self-insured groups. Such pricing issues include the development of service fees using claim closure information, the selection of service durations, and the design of incentive (either activity-based or financially-based) service contracts. Formulas for pricing new and open claims are provided.

Key words and phrases: self-insurance, service length, new claims, open claims

1 Introduction

Self-insurance programs are designed to capture the potential cash flow benefits arising from loss reserves and expense savings. To achieve these goals, self-insured employers and self-insured groups need to carefully select a professional service provider, also known as a third party administrator (TPA). TPAs have substantial experience in claims handling, and they usually have access to other supporting services

---

*Hou-Wen Jeng, A.C.A.S., A.Re., is an actuary with Nationwide Insurance. He holds a B.A. in economics from National Taiwan University and an M.A. in economics from the University of Rochester.

Mr. Jeng’s address is: Nationwide Insurance Co., One Nationwide Plaza, Columbus OH 43216, USA. Internet address: jengh@nationwide.com

†This paper is based on a previous paper entitled “TPA Service Pricing and Incentive Contracts” that appeared in Alternative Markets/Self Insurance—1996 Discussion Paper Program, a nonrefereed publication of the Casualty Actuarial Society. The author thanks Rich Cundy, Rudy Palenik, Virginia Price, the editor and the anonymous referees for their comments. All remaining errors are his responsibility.
such as actuarial, loss control, managed care, and return-to-work programs. Thus, a TPA generally is regarded as the centerpiece of many self-insurance programs.

From a service standpoint, the most significant difference between a TPA and a claims department of an insurance company is that a TPA provides claims services with a variety of service lengths, ranging from twelve months to the life of the claim. The primary product lines for self-insurance are workers' compensation and general liability which are also considered long-tail lines in insurance. Long-tail claims not only complicate the pricing for TPAs, but greatly affect the TPA fee options and service lengths available on the market. Given the long-tail nature of the product lines and the variety of the service lengths, TPAs in general have difficulties in forecasting the costs and pricing their products.

Techniques used in insurance ratemaking and reserving may shed some light on how TPA service pricing should be performed. The aggregate approach used in insurance regarding unallocated loss adjustment expenses (ULAE), however, is not appropriate for pricing TPA products. A more detailed approach using service time and closing ratio by claim age works well in predicting claim handling costs for various service lengths. Here we emphasize the significance of using claim age in the service fee development. Service level is assumed to be a function of claim age. The distribution of claim ages is related to claims closure distributions. This paper illustrates how information can be combined in the development process.

The last pricing issue discussed is the design of incentive contracts. This has become increasingly important for TPA pricing, especially in financial incentive contracts, due to surging market demand. Two major types of performance measurements for incentive contracts are discussed, and a recommendation is made considering factors that impact the financial results of a self insurance program.

When discussing TPA pricing procedures and incentive contracts, the paper focuses on workers' compensation. The formulas, procedures, and examples can be generalized to include other lines such as general liability and auto liability.

2 Fee Options and Service Length

TPA service pricing is not examined as closely by state regulatory agencies as is insurance pricing. State agencies may assume that, like reinsurance pricing, both parties are large and knowledgeable regarding
As a result, the pricing of TPA service contracts is extremely competitive and TPAs usually customize their products to fit the needs of clients.

A TPA typically is expected to provide several service fee options, including per claim, dedicated office/unit, percent of incurred, and percent of paid. There may be one or more choices of service length for each of the fee options, ranging from 12 months to the life of the claim. Table 1 lists the major TPA service fee options and the service lengths available for the corresponding fee option.

### Table 1: Major TPA Service Fee Options

<table>
<thead>
<tr>
<th>Fee Options</th>
<th>Service Length Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Claim</td>
<td>12 months</td>
</tr>
<tr>
<td></td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Life of partnership</td>
</tr>
<tr>
<td></td>
<td>Life of claim</td>
</tr>
<tr>
<td>Dedicated Office/Unit</td>
<td>Same as contract period</td>
</tr>
<tr>
<td>% of Incurred/Paid Loss</td>
<td>Usually life of claim</td>
</tr>
<tr>
<td>% of Premium</td>
<td>Usually life of claim</td>
</tr>
<tr>
<td>% of Employees</td>
<td>Usually life of claim</td>
</tr>
</tbody>
</table>

Before elaborating on the different TPA service fee options, it is helpful to discuss several pricing-related risks: claim frequency, claim severity, loss development, and premium adequacy. Per claim involves two risks: claim severity and loss development. TPAs can mitigate the risks by charging different fees for different types of claims if such a classification is feasible. For example, in workers' compensation, the service fees for medical-only and indemnity claims are significantly different.

If the service fee is based on a percentage of premium, then the TPA faces three risks: frequency, severity, and premium adequacy risks. In other words, the TPA has to absorb the servicing cost of any unexpected increases in claim frequency and severity, and needs to make sure that

---

1. Very rarely do TPAs charge by hours of time, especially for long tail lines because it is difficult to pinpoint the exact time spent for each claim.
2. Here we only consider the pace of the loss development.
the calculated premium is adequate. On the other hand, if the service fee is based on a percentage of paid/incurred losses, the TPA needs to be concerned with the pace of the loss development (although the risks of frequency and severity may be smaller than the premium-based pricing). In practice, many TPAs wish to assume as little insurance-related risks as possible. The fear is that they may be considered as insurers and be regulated as such. In addition, they may not have the resources and insurance expertise to underwrite insurance risks.

2.1 The Per Claim Fee Option

2.1.1 Basics

Because it is flexible in service length, the per claim fee option has been the most popular choice among self-insureds, where service fees are based on the number of claims received by the TPA in the contract period. Under per claim, a self-insured client can choose from various service lengths for the claims to be serviced, such as 12 month, 24 month, life of partnership, and life of claim. This diversity in service length contrasts with traditional insurance where insurers always service claims to conclusion.

The fee for the 12 month claims service provides claims handling on new claims reported during the contract period and claims open at the beginning of the contract period for a period of 12 consecutive months. Similarly, the 24 month claims service provides claims handling for 24 consecutive months. Consider an example where the contract period is from 1/1/95 to 12/31/95 and 24 months is the selected service length. A claim reported on 3/1/95 will be serviced continuously until 2/28/97, 14 months after the end of the contract period. Similarly, a claim reported on 7/20/95 will be serviced continuously until 7/19/97. The total fee calculation using the data in Table 2 is simple:

Total Fee Charges on 12/31/95 = $250 \times 200 + $550 \times 300 = $215,000.

For a new customer, the charges for the open claims assumed at the inception of the contract can be easily determined and billed. New claims (i.e., claims that have never been serviced by a claims administrator) are billed only when they are reported to the TPA. As a result,

3 Most self-insureds report their payrolls and incurred losses to the state. They do not calculate their premium, and their exposures usually are not properly classified as is required in insured cases.

4 See Section 2.3 for a more detailed discussion.
Table 2

Data for a New Customer

<table>
<thead>
<tr>
<th>Contract Period</th>
<th>CY 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Length</td>
<td>24 months</td>
</tr>
<tr>
<td>Per Open Claim Charge</td>
<td>$250</td>
</tr>
<tr>
<td>Per New Claim Charge</td>
<td>$550</td>
</tr>
<tr>
<td>Number of Open Claims Known as of 1/1/95</td>
<td>200</td>
</tr>
<tr>
<td>Number of New Claims During Contract Period</td>
<td>300</td>
</tr>
</tbody>
</table>

CY Z = Calendar Year Z = 1/1/Z - 12/31/Z.

the total service charges under per claim cannot be determined until the end of the contract period. The billing process can become complicated when a customer chooses different service lengths from contract to contract. Consider the following per claim contracts (given in Table 3) for a new customer starting in 1995:

Table 3
Sample Contracts

<table>
<thead>
<tr>
<th>Contract Period</th>
<th>Contract 1</th>
<th>Contract 2</th>
<th>Contract 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Length</td>
<td>24 months</td>
<td>12 months</td>
<td>12 months</td>
</tr>
</tbody>
</table>

CY Z = Calendar Year Z = 1/1/Z - 12/31/Z.

Contract 1 and contract 2 have different service lengths. New claims reported in 1996 and 1995 will be billed as open claims in 1997 if they remain open on their first and second anniversary dates, respectively. In addition, all the open claims assumed at the inception of contract 1 will be billed again if they are open on 1/1/97. Because the service length for contract 3 is 12 months, they will be available for billing again on 1/1/98 if they are not closed by then.

To make the per claim billing process more complicated, a client can choose different service lengths for new and open claims by location and contract. To ensure receiving proper credits, TPA billing must be claim-specific and should track the status of individual claims including service length and claim anniversary date. In practice, if the current contract is not renewed, it is common for TPAs to cease servicing all
claims at the end the current contract period unless the service length is life of claim.\textsuperscript{5}

Life of claim services provide claims handling until settlement at a fixed cost for new and open claims reported to the TPA during the contract period. Life of partnership services are the same as life of claim services except the TPA will stop servicing all existing open claims if the contractual relationship between the TPA and the self-insured ceases. Due to competitive pressure, some TPAs may sell life of partnership service under the guise of life of claim service with a lower price to gain customers. Self-insureds should study the language of their service contracts regarding service length to avoid the consequences of this confusion.

2.1.2 Issues

Self-insureds can reduce claims-servicing cost by choosing a service length that best fits their self-insurance program. For example, if a self-insured finds that most of its claims can be closed within two years, a 24 month service plan may be the best choice. A tail claim service\textsuperscript{6} can be purchased to handle any remaining open claims after two years of service. On the other hand, from a TPA's perspective, the longer the service length, the more uncertainty in service pricing and revenue accrual. Thus, to avoid adverse selection, a TPA needs to determine appropriate pricing relativities between different service lengths, to investigate the closure patterns of prospective clients, and to impose risk charges for longer service lengths.

Similar to unearned premium reserves in insurance, portions of the TPA revenue from a service contract need to be deferred when the service length runs across two or more calendar years. The straight-line method used in calculating unearned premium reserves cannot be applied to the calculation of TPA service fee deferrals because of the uneven service levels at the various development ages of a long-tail claim.

\textsuperscript{5}Surprisingly, there is rarely a fee adjustment for services that have not been performed. The reason may be that most self-insureds do so voluntarily. In workers' compensation, using two TPAs (one for existing open claims and one for new claims from the same work site) may cause significant confusion. On the other hand, it would not be in the interest of the self-insured to have the same TPA handle their existing open claims due to a lack of financial incentives on the part of the TPA. It is a windfall to the TPA as fewer services need to be performed. I believe the financial effect is not significant, however, because the majority of the cancellations are 12 month service contracts. Nevertheless, one can easily factor in this effect in the pricing formula using a historical cancellation rate.

\textsuperscript{6}A tail claim service handles all remaining open claims to conclusion.
As a general rule, the older the claims, the less time they need for service. But a more relevant question to ask is how much of the service fees should be deferred? To answer the question, one must know the claim closure distribution and the average amount of time examiners spend on the claim. The pricing procedure discussed below uses this information in determining service charges for per claim. The deferral percentages can be calculated from this procedure.

For contracts with long service length, casualty actuaries can provide valuable services in the areas of TPA pricing and revenue deferral. Most self-insureds, however, are just as uncomfortable as TPAs in entering a contract with a long service length. In practice, 12 month handling is the predominant choice by self-insureds for their TPA service contracts. This phenomenon can be attributed to the following three reasons:

- Because most self-insureds are generally cost conscious, the selection of a shorter duration service plan can help their cash flow.

- Shorter service durations make it easier for a self-insured to move its program to another TPA if it is not satisfied with the current TPA's services.

- When the service contract for future claims between a TPA and a self-insured is not renewed, it would not be in the interest of the self-insured to have the same TPA handle its existing open claims due to a lack of financial incentives on the part of the TPA. In the case of life of claim handling, the self-insured and the TPA need to be in close contact regarding claims handling for many years after the termination of the service contract.

From a TPA's point of view, a contract with a short service length does have its down side. More components such as the handling of the remaining open claims from prior contracts must be negotiated at the contract renewal, and renewal negotiations occur more frequently. As a result, TPA's overhead expenses may be significantly increased. If the majority of the TPA contracts have short service length, it would be difficult for a TPA to project its future claim volumes and revenues.

### 2.2 Dedicated Office/Unit

Dedicated office/unit is an option where a TPA establishes a claims office or a claims unit to exclusively handle claims for the client. The set-up cost and the subsequent administrative costs, as well as the TPA's overhead and profit, are paid by the self-insured.\(^7\) Under this

---

\(^7\)This resembles mark-up or cost-plus pricing.
option, the service length for all claims, regardless of age, is the same as the contract period. If the contract is not renewed, the TPA will stop servicing all claims at the end of the current contract period. This option poses the least pricing risk to a TPA as it has none of the frequency, severity, loss development, and premium adequacy risks, and expenses are billed as soon as incurred. This option is usually more expensive, however, and is only recommended for larger self-insureds.

To self-insureds, the major advantage of such an arrangement is that claims examiners are familiar with the self-insured and thus are able to satisfy the client's special needs in claims handling. In addition, the location of the dedicated office can be selected strategically so that most of the current and potential claimants can be in the vicinity of the claim office. This is especially beneficial to clients such as municipalities and school districts that are geographically concentrated.

An insurance company theoretically can minimize its total payout by allocating its resources between losses and adjustment expenses. Doing so recognizes that spending more on loss adjustment may reduce loss payments and potentially can result in a lower overall cost because of the better claims management.

By being self-insured and choosing the dedicated office/unit option, a customer controls its resource allocations and is able to dictate the degree of care and the amount of time examiners spend on each case. One can demand more claim examiners to service a fixed number of claims (i.e., a lower caseload per examiner) and thus provide better service to claimants. Others may opt for a higher caseload per examiner to save adjustment expenses. Thus, under dedicated office/unit, the role of the TPA is reduced to providing the staff, computer systems, and other related technical services while the client makes the more important financial decisions and determines the extent of the claims services.

2.3 The Percentage Approach

Based on a predetermined percentage of the base figure (e.g., incurred loss) this fee option includes three major varieties: percent of incurred loss, percent of paid loss, and percent of premium. The service length is usually the life of the claim, as it would be difficult to determine the service fee by claim age. In the case of percent of incurred loss, compared to other pricing options, this option involves little, if any, insurance-related risk. Thus, the pricing risk refers to general business risks such as employees, rents, equipment leases, and so on. For example, when the contract cancels, the TPA still needs to pay the rents and salaries for a certain period of time.
loss, one must determine the pace of the incurred loss development in order to price a 12 month service contract. Such information is usually not available from the self-insured. Percent of premium is used less frequently than the other two, perhaps because this option requires more information and insurance expertise for underwriting.

Both percent of paid loss and percent of incurred loss are highly individualized pricing approaches, where service charges for any claims are directly related to the cost of claims. A TPA must monitor the paid or incurred amount to determine if additional billings are necessary. Consider a claim whose ultimate cost is initially estimated at $30,000. Later it is found that a medical treatment is needed for an additional $20,000. Assuming the TPA fee is set at 7 percent of incurred loss, the fee charge for this claim will increase from $2,100 to $3,500 due to the medical treatment.

From the outset, it appears that both methods are equitable ways to determine compensations for TPA services if the percentage is selected appropriately. A closer look, however, reveals that there are serious drawbacks inherent in the methods. First, the perception of a TPA as an independent third party in claims handling could be lost because TPA service fees are linked to the settlement amount. TPAs may have little incentive to control claim costs. Second, it is also difficult for TPAs to manage the billing because incurred and paid amounts for individual claims change constantly. Third, although for any claims the paid amount eventually equals the incurred amount, the timing of the claim payments under percent of paid dictates how quickly the TPA can bill its clients. For example, most of the claims in litigation are not paid until the legal issues are resolved. At the same time, most of the handling service work on those claims has already been done. Thus, depending upon the underlying frequency and severity distributions, the use of percent of paid may result in significant risk-taking on the part of the TPA in terms of potential cash flow problems.

3 Development of TPA Service Fees

3.1 Insurance Ratemaking and Reserving Considerations

In insurance ratemaking and reserving, unallocated loss adjustment expenses (ULAE) are estimated on an aggregate basis. For example, the provision for ULAE in insurance rates generally is assumed to be a certain percentage of the premium using industry experience. The reserves for ULAE usually are estimated using the ratio of the historical
ULAE to loss and allocated among individual accident years. In the annual statement of insurance companies, the ULAE reserve calculation is based on the assumption that 50 percent of the ULAE is paid when the claim is reported and the other 50 percent is paid when the claim is settled.

There have been few changes in the ways that ULAE is built into rates and how ULAE reserves are calculated. There appears to be no need for insurance companies to establish a higher level of accuracy in the estimation of ULAE. After all, the provision for ULAE accounts for, on average, only 6 percent of the rate, and the variations in loss generally overshadow those in ULAE.

On the other hand, because a TPA's major business is claims handling, the ability to break the cost down by claim type and service length is important to the pricing of TPA services. The aggregate approach and the ad hoc rules used in insurance ratemaking and reserving are inadequate for TPA service pricing.

The following section describes an approach using service time and closing ratio by claim age to predict per claim handling costs for various service lengths. The major assumption is that service level is a function of claim age.9 The objectives are to keep the model simple and to explain most variation in service level. One may argue that service level also depends on other factors, such as the seriousness of the claim. If such factors are also correlated with claim age, however, the assumption has implicitly considered them.

3.2 Per Claim Pricing

3.2.1 New Claims

We will now explore how claims closure and service level information can be used to develop per claim service fees. Service level (i.e., examiner time) is assumed to be a function of claim age.

Let \( t (t = 0, 1, 2, \ldots) \) be claim age measured in months10 and \( F(t) \) be the cumulative percentage of closed claim at the start of the \( t \)-th month, with \( F(0) = F(1) = 0 \). Thus, \( (1 - F(t)) \) can be viewed as the probability that a claim will be open at the beginning of the \( t \)-th month since it was

---

9Claim age or time are not the only factors that should be considered. The emphasis is whether the explanatory factor can be objectively measured and if the related data are readily available for estimation. If the TPA has a consistent claims practice, claim age seems to be a natural choice.

10The analysis that follows can easily be adjusted to deal with claim age measured in other time units such as quarters or years.
reported to the TPA. Let \( g(t) \) be the average service time (measured in hours) spent on an open claim at age \( t \) (i.e., number of hours examiners spend on a case at age \( t \) months.)\(^{11}\)

The shape of \( g(t) \) may take many forms depending on the line of business and the type of claims. Two types of service time curves are usually observed in the case of workers' compensation indemnity claims: a downward sloping curve and a humped curve with its peak within the first six months (see Figure 1). Both curves indicate that most of the service time for an average claim is spent in the first 18 months, which contrasts with the common belief that older claims require more service time per month to settle than those claims that are settled early and quickly.

![Figure 1: A Downward Sloping Curve and a Humped Curve](image)

The two most time-consuming activities of claims adjusting are the investigation of injuries to determine compensability and the coordi-

---

\(^{11}\)In most cases, the information needed to compile \( F(t) \) is readily available from the TPA's computer system. On the other hand, the estimation of \( g(t) \) may involve a detailed study on how claims examiners spend their time on claims with different ages. If such a study is not possible, \( g(t) \) can be determined from the result of a survey based on examiners' experience and judgment.
nation of medical treatments that include surgeries and rehabilitation. Because these activities occur more frequently in the early stage of the claims, \( g(t) \) is usually a downward sloping curve or a humped curve for workers' compensation indemnity claims.\(^{12}\)

The next step is to determine the unit cost of examiner time (including salary, benefits, overhead, and profit) at the beginning of the contract period. For example, assume that the annual salary and benefits for an examiner are given at $50,000 while overhead and profit account for 50 percent of the cost. Given that the total working hours in a year are 2,000 (250 working days and eight hours per working day), the unit cost of examiner time can be set at $50 per hour \([($50,000/0.5)/2,000] \).

Let \( P^{(n)}(k) \) be the price for handling a new claim from month 1 to month \( k \) \( (k = 1, 2, \ldots) \) and let \( c \) be the hourly cost of service time at the beginning of the contract period. Further assume that \( c \) increases at a rate of \((1 + s)\) per month, \( s \geq 0 \). Thus, the hourly cost at the start of the \( t \)-th month is \( c_t = c \times (1 + s)^{t-1} \), for \( t = 1, 2, \ldots \). Then the per claim service price function for a new claim is given by:

\[
P^{(n)}(k) = \sum_{t=1}^{k} c_t \nu^{t-1} g(t) (1 - F(t)) \quad \text{for } k = 1, 2, \ldots \tag{1}
\]

where \( \nu = 1/(1 + i) \) is the monthly interest discount factor, with \( i \) being the monthly interest rate. The discount factor \( \nu \) can be selected by the TPA to reflect its cost of capital and other needs. We assume that all service time is rendered at the beginning of every month and, thus, discounting takes place at the beginning of each month, i.e., at time \( t - 1 \).

Equation (1) can be rewritten as

\[
P^{(n)}(k) = c \sum_{t=1}^{k} \beta^{t-1} g(t) (1 - F(t)) \tag{2}
\]

where \( \beta = (1 + s) \nu \). Thus the per claim service price for a new claim to be handled to settlement (for life)\(^{13}\) is \( P^{(n)}(\infty) \), while that for new claim service price for 12 month handling is \( P^{(n)}(12) \), and so on.

\(^{12}\)In establishing \( g(t) \), a TPA needs to consider segregating its experience into more homogeneous groupings. Experience may be subdivided by claim type or location (i.e., service time may be different as required by regulation. California and Texas are good examples.)

\(^{13}\)In this case, to ensure that \( P^{(n)}(\infty) \) is finite, we must have \( \beta < 1 \) or have the maximum number of years that it can take to settle a claim be bounded. In practice, the latter condition is not restrictive because one can expect all claims to be settled within say 30 years or 50 years or even 100 years.
Table 4 shows how service time and claim closure information are combined to develop the service fees for per claim. The cumulative closing percentage \( F(t) \) in Column 4 at the beginning of the first month \( t = 1 \) is zero. By the end of the month, 10 percent of the claims are closed and the service time rendered in the month is ten hours per claim. Thus, the expected service time for the first month is ten hours as indicated in the last column of the table. \( g(t) \) is the service time for each claim open at age \( t \). For the second month, \( g(2) \) is 14 hours and \( (1 - F(2)) \) is 90 percent. Therefore, the expected service time in the second month is 12.6 hours. It is straightforward to calculate the expected service time for the remaining months. Thus, for example, equation (2) leads to

\[
P(n)(\infty) = c[10 + 12.6\beta + 11.4\beta^2 + \cdots + 0.78\beta^{11} + \cdots 0.1\beta^{23} + \cdots].
\]

<table>
<thead>
<tr>
<th>( t )</th>
<th>( g(t) )</th>
<th>Closing %</th>
<th>( F(t) )</th>
<th>( 1 - F(t) )</th>
<th>( g(t)(1 - F(t)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 hours</td>
<td>10%</td>
<td>0%</td>
<td>100%</td>
<td>10 hours</td>
</tr>
<tr>
<td>2</td>
<td>14 hours</td>
<td>14%</td>
<td>10%</td>
<td>90%</td>
<td>12.6 hours</td>
</tr>
<tr>
<td>3</td>
<td>15 hours</td>
<td>12%</td>
<td>24%</td>
<td>76%</td>
<td>11.4 hours</td>
</tr>
<tr>
<td>4</td>
<td>13 hours</td>
<td>11%</td>
<td>36%</td>
<td>64%</td>
<td>8.32 hours</td>
</tr>
<tr>
<td>5</td>
<td>10 hours</td>
<td>10%</td>
<td>47%</td>
<td>53%</td>
<td>5.3 hours</td>
</tr>
<tr>
<td>6</td>
<td>8 hours</td>
<td>9%</td>
<td>57%</td>
<td>43%</td>
<td>3.44 hours</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>3 hours</td>
<td>2.5%</td>
<td>74%</td>
<td>26%</td>
<td>0.78 hours</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>24</td>
<td>1 hour</td>
<td>1.2%</td>
<td>90%</td>
<td>10%</td>
<td>0.1 hour</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

For life of partnership handling, a subjective probability distribution has to be included to indicate the possibility of cancellation. In general, it is assumed that the average time of a contractual relationship between a self-insured and a TPA is three to five years. Consequently, the variations in life of partnership pricing among TPAs can be significant, depending critically on the expectation and the risk tolerance level of the TPA.
In establishing claim closure distributions, a TPA needs to consider segregating its experience into more homogeneous groupings. Long-tail lines usually exhibit distinctive closing patterns compared to other product lines. Even within a long-tail line it is usually beneficial to subdivide experience by claim type. For example, in workers' compensation, most medical-only claims can be closed within six months while some indemnity claims can linger for more than five years.

There is no doubt that this procedure can establish only a baseline for pricing while much of the pricing decision has to be based on the underwriting characteristics of the customers. One needs to examine, among other things, the claim closing patterns of the prospective clients in order to determine the deviation of their experience from the TPA's own experience and adjust the price accordingly.

### 3.2.2 Open Claims

Let $P^{(o)}(m, n)$ be the service fee for an open claim at age $m$ to age $n$, for $m = 1, 2, \ldots$ and $n = m + 1, m + 2, \ldots$. Using the same notations as in Section 3.2.1, $P^{(o)}(m, n)$ can be calculated as follows:

$$P^{(o)}(m, n) = \sum_{t=m}^{n} c_t v^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))}$$

$$= c (1 + s)^{m-1} \sum_{t=m}^{n} \beta^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))}. \quad (3)$$

In practice, service charges for claims open more than 12 months are seldom based on individual claim age, as it would be tedious to calculate the fees. A weighted-average charge is applied to each open claim regardless of its age. Assuming the claim volume from year to year is stable, the formula for the weighted average charges is:

$$P^{(o)}_w(m, n) = c (1 + s)^{m-1} \sum_{j=12}^{\infty} w_j \sum_{t=m}^{n+j} \beta^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))} \quad (4)$$

where

$$w_j = \frac{1 - F(j)}{\sum_{k=12}^{\infty} (1 - F(k))}$$

is the probability weight used for the $j$-th month.

---

There is no assumption of universal efficiency of all TPAs. The poor claim experience of a customer, for example, when compared to similar risks, may result from the random variations of claims, the poor management of its TPA, or simply reflect the customer's true exposures.
3.3 State-Group Relativities for Per Claim

For a TPA with clients in multiple states, there is a need to differentiate service costs among states. To calculate per claim charges by state, one can establish state-group relativities similar to those used in class ratemaking in property/casualty insurance pricing. Once state-group relativities are established, updates of the base price for each state can be performed easily.

The criteria to divide states into state-groups with similar claims handling costs can be based on the TPA's internal claims closure experience and cost by state, supplemented by statistics from national or state rating bureaus. For workers' compensation, important statistics include the percentage of serious cases and the per claim severity which may differ significantly by state. In addition, the degree of state regulation which is always an important contributing factor to TPA's service costs can help determine the makeup of the state-groups.

Specific actuarial techniques and much more data are needed to establish credible estimates of the values of state-group relativities. Even a national TPA may not have enough information in all claim categories for all states. For local or regional TPAs, state-group relativities can be set only judgmentally based on the TPA's internal cost and published information from state rating bureaus.

4 Incentive Contracts

The last pricing issue to be discussed is the design of incentive contracts. There has been a strong interest among self-insureds to establish a relationship between service fees and TPA performance in order to monitor the effectiveness of TPAs in controlling claim costs. Essentially, an incentive program requires that a certain percentage of the service fees be set aside for a bonus or penalty based on several performance measurements of the TPA services. The results of the performance measurements valued as of predetermined dates are compared to negotiated targets for the calculation of the bonus or penalty.

Before discussing any specific performance measurements, it is useful to set some common sense criteria to evaluate their feasibility. The following provides a reasonable checklist for such purposes:

- The TPA has sufficient control over the performance measurement;
- The value of the performance measurement can be objectively determined, and both parties have the ability to track results; and
There exist reliable benchmark data for comparison.

4.1 Basics

There are two major types of performance measurements: activity-based measurements and financial measurements. Popular measurements of TPA performance are usually activity-based such as number of claims closed by age, timely bill payments, timely claim processing, and reserving adequacy. The usual financial measurements for incentive programs include paid loss and incurred loss.

Most activity-based measurements can satisfy the three criteria. Take timely bill payments and claim processing as examples. An incentive program can stipulate that claim bills should be paid by the TPA within two business days after receiving the bills, or that claimants should be contacted within 24 hours after the claim is reported. The data for calculating such performance measurements should be available from the TPA's system and the results of the measurements can be determined easily. Therefore, the implementation of such an activity-based incentive program is usually straightforward.

4.2 Financial Incentive Contracts

The TPA industry has been experiencing more demand for financially based measurements, such as comparing actual and target incurred/paid amounts for claims incurred within the service contract period. In general, TPAs are hesitant to accept financially based measurements as they may appear to be taking insurance risk in which they have insufficient knowledge and little interest. Given that financial-incentive contracts have gained considerable popularity in recent years, the TPA industry has been forced to develop measurements that are mutually agreeable to the claims administrator and the self-insured.

Total policy year paid or incurred loss by development age have been suggested as performance measures for a risk-sharing program. Paid or incurred loss by development age is measured against an index such as policy year payroll before it is compared to a predetermined goal. Using the criteria described at the beginning of this section, it is clear that the amount of paid or incurred loss by development age per se can be determined easily. The TPA does not have sufficient control over the measures, however, as any total losses are affected by frequency, exposure, inflation, and other factors. In addition to the volatility of paid and incurred losses, it is difficult to find reliable data for benchmarking. Although these drawbacks may seem obvious to casualty actuaries,
many self-insureds insist on using changes in paid-to-date or incurred-to-date loss as performance measurements.

4.3 A Suggestion: Use Averages

Take workers' compensation as an example. There are four factors that can significantly change the financial results of a self-insured program: exposure (payroll) changes, state benefit changes, claim frequency changes, and inflation. A TPA should not be responsible for variations due to changes in exposure, frequency, and benefit level because none of these factors can be controlled by a claims administrator. For example, higher frequency in reported workers' compensation claims can be the result of a lay-off, which is beyond the control of the TPA.

To eliminate the impact of frequency changes on total loss, it seems appropriate and equitable to use incurred per claim severity as a performance measure for a financial risk-sharing plan. By eliminating the variations in frequency and exposure, per claim severity usually exhibits stable development patterns, given sufficiently large claim volumes. Most importantly, per claim severity can be managed and partially controlled by the TPA. Thus, it appears to be an ideal candidate for measuring TPA performance.

Additional benefits of using per claim severity as a performance measure are:

- There is no need to compare per claim severity to payroll or number of employees for incentive contract purposes;
- The industry average cost per claim by state is available from state rating bureaus;\(^{15}\) consequently, benchmarking should be easier and the results should be much more reliable;
- By comparing to an industry average, the variations due to changes in benefit level can be eliminated.

Per claim severity should be used on an ultimate basis as a performance measurement.\(^{16}\) Only when the baseline for comparison is established on an ultimate basis can the loss experience of a policy year be evaluated. The results can be misleading if one is merely looking

---

\(^{15}\) Precautions must be taken when bureau data are used, as trends and development may be needed.

\(^{16}\) A method for estimating ultimate values needs to be agreed upon in advance to avoid competing estimates from the TPA and the self-insured.
for the incremental changes between two development ages that are subject to the timing of claim payments and reserve recording.

An incentive contract ideally can look and operate in a way similar to retrospective rating plans. To establish a baseline for a policy year, the usual actuarial methods including capping large losses can be applied to loss data in the estimation of the ultimate severity. This can be done six months after the end of the policy year, the same time when retrospective rating plans start to evaluate policy year experience. The main difference is that in retrospective rating the target incurred loss is revised every 12 months thereafter until the final settlement of the policy year, while in incentive contracts a baseline (i.e., estimated ultimate severity) is determined six months after the end of the policy year for benchmarking purposes at later dates. A bonus or penalty can be calculated based on the deviation of the projected ultimate per claim severity at a later evaluation date (e.g., 30 months after policy inception) from the baseline. A subsequent computation/adjustment can be performed every 12 months until both parties agree that the latest computation will be the final one for the policy year.

5 Concluding Remarks

One important component that is missing in TPA pricing is self-insurance database support. Self-insured entities do not report loss, payroll, or other relevant experience data to state rating bureaus. To meet their pricing needs, TPAs rely on their own experience or purchase data from state rating bureaus, which may or may not be appropriate for the self-insurance purposes. The National Council on Compensation Insurance has initiated a program for collecting loss data on self-insured groups. This may be a good start toward a more complete and reliable database for TPA pricing.

With the introduction of managed care organizations (MCOs) in many states, the role of TPAs in the business of claims handling may fundamentally change. Judging from developments over the past few years, TPAs and MCOs may have to share the responsibilities in medical cost containment, rehabilitation, and return-to-work programs. On the other hand, TPAs may be in an excellent position to launch their own medical networks and merge these two functions. It will be interesting to see how these changes will impact the pricing of traditional TPA services and the expanded services provided jointly by a TPA and an MCO.
References


Annuity Choices for Pensioners
M. Zaki Khorasanee*

Abstract†

We consider two ways for a retiree to obtain a pension from a retirement fund: through the purchase of a whole life annuity providing a level monetary income; and through the withdrawal of income from a fund invested in equities. Deterministic and stochastic models are used to assess the risks and benefits associated with each approach. In each case the projected cash flows are compared with those from a whole life annuity providing an income linked to price inflation. We conclude that, although each of the two options considered involves significant risks, each method may be attractive to certain groups of pensioners, particularly those with additional savings held outside the retirement fund.

Key words and phrases: index-linked, inflation, equity portfolio, sinking fund, break-even duration

1 Introduction

In the United Kingdom (U.K.) certain types of pension plans provide a lump sum benefit at retirement rather than monthly (or other periodic) payments.1 The manner in which the lump sum is invested

---

* M. Zaki Khorasanee, B.A., F.I.A., is a lecturer in the Department of Actuarial Science and Statistics at City University, England. He obtained his B.A. degree in 1985 from Cambridge University and has worked for six years as a pension actuary with various consulting firms in the London area.

Mr. Khorasanee's address is: Department of Actuarial Science and Statistics, City University, Northampton Square, London EC1V 0HB, England. Internet address: m.z.khorasanee@city.ac.uk

† The author would like to acknowledge the helpful comments made on earlier drafts of this paper by his colleague Professor Steven Haberman and by the referees.

1 These include individual insured plans, defined contribution plans, and defined benefit plans in which the benefit formula is for a lump sum at retirement.
is restricted by legislation. A retirement fund must be established to provide income although a portion of the fund, usually not exceeding 25 percent, may be taken in cash. The safest option may be to buy a whole life annuity providing an income guaranteed to increase with an index of consumer price inflation. In the U.K. such a product is called an *index-linked annuity*.²

This paper considers the merits of two alternatives to the index-linked annuity that currently are available in the U.K.:

- A level whole life annuity providing a stable income, the most popular option; and
- An income withdrawal option in which no annuity is purchased, but the pensioner draws an income from the retirement fund.

We examine the implications of each by comparing the projected cash flows against those from an index-linked annuity.

### 1.1 Why Not Choose an Index-Linked Annuity?

Index-linked annuities are relatively unpopular with U.K. consumers because the income obtained is initially much lower than the income provided by a level annuity. The following table shows the initial income available from the most competitive U.K. insurer in July 1996 for each type of annuity.³

In July 1996 consumer price inflation was approximately 3 percent per annum in the U.K. If this rate of inflation were to continue, an index-linked annuity would provide a lower income for 13 years from age 60, 11 years from age 65, and 10 years from age 70. Pensioners may believe that even if they live a reasonable life span beyond these durations, they probably would receive a higher aggregate income from a level annuity, and we show that this opinion is justified.⁴

The purchase of a level annuity has two potential disadvantages:

- The pensioner may live well beyond his/her life expectancy;

---

² The index used is the Retail Prices Index. It is the most widely used measure of U.K. consumer price inflation and is published monthly by the government.

³ Retiring members are normally given an open market option, i.e., the right to buy an annuity with the insurance company of their choice. As a consequence, most U.K. life annuity business is written by a relatively small number of insurers offering the most competitive rates.

⁴ The apparently poor rates offered for index-linked annuities may be a consequence of mortality selection—lives in poor health are unlikely to opt for index-linked annuities. In analyzing the choice for any one individual, however, we assume the mortality of that individual will be the same whichever type of annuity is purchased.
Khorasanee: Annuity Choices for Pensioners

Table 1
Annual Income to a Male Pensioner From a £10,000 Fund

<table>
<thead>
<tr>
<th>Retirement Age</th>
<th>Level Annuity</th>
<th>Index-Linked Annuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>£1,016</td>
<td>£683</td>
</tr>
<tr>
<td>65</td>
<td>£1,119</td>
<td>£797</td>
</tr>
<tr>
<td>70</td>
<td>£1,280</td>
<td>£951</td>
</tr>
</tbody>
</table>

Source: Annuity Direct (July 19, 1996), available from Ceefax, a teletext service of the British Broadcasting Corporation (BBC).

- Inflation may be higher than anticipated.

For these reason, purchasers of level annuities are sometimes advised to save a substantial part of the income they obtain in the early years to build a fund that can provide some protection against inflation and longevity. In a later section we examine the efficacy of such a policy by assuming that pensioners save (or dis-save) the difference between the income from their level annuity and the income that would have been obtained from an index-linked annuity. Even such a conservative policy does not immunize the pensioner against risk.

1.2 The Income Withdrawal Option

The income withdrawal option was introduced in the U.K. in 1995. This option allows the pensioner to defer the purchase of an annuity, instead the pensioner draws an income by utilizing assets in the fund. Certain restrictions apply, however: the income drawn must lie between 35 percent and 100 percent of the income that otherwise could have been obtained from a level annuity, and the pensioner cannot delay the purchase of a whole life annuity beyond age 75.

The income withdrawal option has two principal attractions:

- At the time of death, the capital remaining in the fund remains part of the pensioner's estate; and

- The fund can be invested in assets that may provide a higher return to the pensioner than available from a whole life annuity.

The first characteristic is sometimes described as capital protection. In a whole life annuity, the absence of a death payout allows a higher
level of income for the living. Thus, the income from a whole life an-
nuity fund will exceed that from an income withdrawal fund if both are
invested in the same assets and are subject to the same expenses. This
difference increases as the pensioner ages because of the increasing
mortality strain.

It follows that an income withdrawal fund must earn a higher return
than a whole life annuity fund if the pensioner is to maintain an equiv-
alent income. In practice, this means that an income withdrawal fund
is likely to be invested wholly or partly in equities, as equities are ex-
pected to outperform the government bonds held by insurers' annuity
funds. The market price volatility associated with equities, however,
creates additional risks for the pensioner.

2 Level VS. Index-Linked Annuities

2.1 Assets of Whole Life Annuity Fund

The income received from a whole life annuity policy depends on
the interest, mortality, and expenses assumed by the insurer for its an-
nuity portfolio. The interest assumption depends on the assets held by
the insurer. In the case of level annuities, the insurer normally holds
a fixed interest bond portfolio roughly matching the mean term of its
liabilities. In the case of index-linked annuities, the insurer normally
holds index-linked bonds, providing interest and principal payments
that increase in step with consumer price inflation. Both types of secu-
rities have been issued by the British government to finance its national
debt, although the total market value of fixed interest bonds currently
in issue is roughly eight times as great as that of index-linked bonds. 5

The prospective return on a fixed interest bond is measured by its
gross redemption yield. 6 The prospective return offered by an index-
linked bond is measured relative to future price inflation and is termed
the real gross redemption yield. This real yield can be thought of as
the interest rate at which the present value of future income and cap-
ital payments from the bond would equal its current market value if
future inflation were zero. The interest rates used to price level and
index-linked annuities closely follow the average yields available on
fixed interest and index-linked bonds, respectively.

5Total market capitalizations were £217bn ($337bn) for fixed interest bonds and
£26bn ($41bn) for index-linked bonds (The Times of London, July 29, 1996, daily busi-
ness section).

6The gross redemption yield is the interest rate at which the present value of future
income and capital payments equals the market value of the bond.
We now define the annual price inflation and other annual rates:

\[ j = \text{Interest rate used to price level annuities;} \]
\[ c = \text{Expected rate of price inflation; and} \]
\[ r = \text{Real interest rate used to price index-linked annuities.} \]

In a scenario of constant interest rates and constant price inflation, we would expect the total return on fixed interest and index-linked bonds to be identical, thus:

\[ (1 + j) = (1 + r)(1 + c). \] (1)

Throughout this section we assume that the difference between the nominal yield on fixed interest bonds and the real yield on index-linked bonds gives an unbiased estimate of future price inflation.

2.2 Deterministic Comparison

We now compare the projected income from both types of annuities assuming a constant force of price inflation. The following notation will be used:

\[ t = \text{Time since the purchase of a level annuity, } t \geq 0; \]
\[ x = \text{The age of a pensioner at the moment of retirement;} \]
\[ \ddot{a}_x^c = \text{Purchase price of an indexed-linked annuity of 1 per annum payable to a pensioner age } x \text{ at retirement;} \]
\[ \int_0^\infty (1 + c)^t (1 + j)^{-t} t p_x \, dt = \int_0^\infty (1 + r)^{-t} t p_x \, dt \]
\[ \ddot{a}_x^l = \text{Purchase price of a continuous level annuity of 1 per annum payable to a pensioner age } x \text{ at retirement;} \]
\[ \int_0^\infty (1 + j)^{-t} t p_x \, dt \]
\[ (AVS)_t = \text{Accumulated value of sinking fund } t \text{ years after the purchase of a level annuity.} \]

We assume that, to maintain his/her standard of living, the purchaser of a level annuity actually spends (consumes) money at a rate
equal to the income generated from an index-linked annuity, the difference being saved in (or withdrawn from) a sinking fund. Thus if a retirement fund of one monetary unit is used to buy a level annuity providing a continuous income, the rate at which income is saved in (or withdrawn from) the sinking fund at time \( t \) is given by \( S_t(j, c) \):

\[
S_t(j, c) = \frac{1}{\bar{a}_x^j} - \frac{(1 + c)^t}{\bar{a}_x^r}.
\]

The present value of the sinking fund accumulation up to time \( t \), discounted back to the purchase date at the nominal interest rate, \( j \), is:

\[
(PVS)_t(j, c) = \int_0^t S_\tau(j, c) (1 + j)^{-\tau} d\tau
= \frac{\bar{a}_{\bar{t}}^j}{\bar{a}_x^j} - \frac{\bar{a}_x^r}{\bar{a}_x^r}.
\]

Initially the sinking fund grows because the money spent by the pensioner is less than the income from the level annuity. As the payment from the index-linked annuity increases continuously at a constant rate, it will eventually exceed the level annuity payments.\(^7\) Thus if the pensioner lives for a sufficiently long time, the sinking fund will become zero and the pensioner will have to withdraw money from the sinking fund.

Let \( N \), called the break-even duration, denote the first time the sinking fund falls to zero. Thus the level annuity provides a more valuable overall income for a pensioner who dies before \( N \), and the index-linked annuity provides a more valuable overall income for a pensioner who survives beyond \( N \). As \((PVS)_N(j, c) = 0\), equation (2) yields

\[
\frac{\bar{a}_N^j}{\bar{a}_N^r} = \frac{\bar{a}_x^j}{\bar{a}_x^r}.
\]

To price these annuities, let us assume \( j = 7 \) percent net of all expenses; \( r = 3 \) percent net of all expenses; \( c = 1.07/1.03 - 1 = 3.88 \) percent; and mortality following the male PA(90) life table.\(^8\) Using standard numerical methods, equation (3) can be solved for \( N \). The results

---

\(^7\)In fact the index-linked annuity payment ultimately is unbounded.

\(^8\)The PA(90) life table is based on the mortality experience of pensioners in employersponsored plans administered by U.K. insurance companies. The table is published by the Institute and Faculty of Actuaries.
are shown in Table 2. In addition Table 2 shows the life expectancy at retirement ($\hat{e}_x$), $N$, and the probability of surviving to age $x + N$, for different values of $x$. Notice that the sinking fund is exhausted one to three years after the pensioner's life expectancy at retirement. The fifth column of Table 2 shows that a retiree has less than a 50 percent chance of surviving to the age $x + N$, and this probability decreases with the age at retirement. A pensioner probably will obtain more value from a level annuity than from an index-linked annuity. The risk of a dramatic fall in income on surviving beyond the break-even duration is significant at all ages. If a pensioner who bought a level annuity at 65 survived to the break-even duration, his or her rate of income at the break-even duration would be reduced by a factor of $(1 + c)^{-N} = 0.514$.

### Table 2

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\hat{e}_x$</th>
<th>$N$</th>
<th>$NP_x$</th>
<th>$(1 + c)^{-N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>26.1</td>
<td>27.8</td>
<td>0.465</td>
<td>0.347</td>
</tr>
<tr>
<td>55</td>
<td>21.8</td>
<td>24.1</td>
<td>0.433</td>
<td>0.399</td>
</tr>
<tr>
<td>60</td>
<td>18.0</td>
<td>20.7</td>
<td>0.401</td>
<td>0.454</td>
</tr>
<tr>
<td>65</td>
<td>14.6</td>
<td>17.5</td>
<td>0.368</td>
<td>0.514</td>
</tr>
<tr>
<td>70</td>
<td>11.6</td>
<td>14.5</td>
<td>0.342</td>
<td>0.576</td>
</tr>
<tr>
<td>75</td>
<td>9.0</td>
<td>11.8</td>
<td>0.303</td>
<td>0.638</td>
</tr>
</tbody>
</table>

2.3 Effect of Uncertain Inflation

We now use a simple model to investigate the effect of uncertainty in the rate of inflation on our previous results. Let us again assume a constant inflation rate, but one that is different from the expected rate, i.e., let $b$ denote the actual rate of price inflation. Using the actual inflation rate, equation (2) is modified as follows:

$$ S_t(j, b) = \frac{1}{\hat{a}_x^t} - (1 + b)^t $$

$$ (PVS)_t(j, b) = \int_0^t S_\tau(j, b)(1 + j)^{-\tau}d\tau $$

---

9By which we mean that the pensioner could reproduce the income from an index-linked annuity and still have a positive sinking fund at death.
where $r' = (1 + j)/(1 + b) - 1$ is the actual real rate of interest. Again we can solve the equation \((PVS)N(j, b) = 0\) to obtain the break-even duration.

More realistically, $b$ will not be known at $t = 0$ because the actual rate of inflation is a random variable. Let $B$ be the random variable denoting the actual rate of inflation, and assume $B$ has a known discrete\(^{10}\) distribution $f_i = \Pr[B = b_i]$, for $i = \ldots, -2, -1, 0, 1, 2, \ldots$. Given that $B = b$, let $N(b)$ be the resulting break-even duration calculated according to the equation

\[
(PVS)_{N(b)}(j, b) = 0.
\]

There are two quantities of interest to us:

\[
E[N(B)] = \text{The expected break-even duration;} \quad \pi_x = \text{Probability of a pensioner who retires at age $x$ survives beyond exhaustion of the sinking fund;}
\]

\[
= \sum_{i=-\infty}^{\infty} N(b_i) f_i \quad \text{and} \quad = \sum_{i=-\infty}^{\infty} N(b_i) p_x f_i. \tag{6} \tag{7}
\]

As an example, consider the following case: $x = 65$, $j = 7$ percent, $c = 3.88$ percent, $B = c + 1$ percent with probability 0.25, $B = c$ with probability 0.50, and $B = c - 1$ percent with probability 0.25. Using each of the three possible values of $B$, equation (5) yields the information in Table 3.

\[
E[N(B)] = 26.0 \times 0.25 + 17.5 \times 0.50 + 13.3 \times 0.25 = 18.575 \quad \pi_{65} = 26.0 p_{65} \times 0.25 + 17.5 p_{65} \times 0.50 + 13.3 p_{65} \times 0.25 = 0.093 \times 0.25 + 0.368 \times 0.50 + 0.543 \times 0.25 = 0.343.
\]

This probability of 0.343 is lower than the corresponding figure in Table 2, for two reasons:

\(^{10}\)In practice, $B$ has a continuous distribution.
Table 3

<table>
<thead>
<tr>
<th>$b$</th>
<th>$Pr[B = b]$</th>
<th>$N(b)$</th>
<th>$N(b)p_{65}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.88%</td>
<td>0.25</td>
<td>26.0</td>
<td>0.093</td>
</tr>
<tr>
<td>3.88%</td>
<td>0.50</td>
<td>17.5</td>
<td>0.368</td>
</tr>
<tr>
<td>4.88%</td>
<td>0.25</td>
<td>13.3</td>
<td>0.543</td>
</tr>
</tbody>
</table>

- The break-even duration increases in the low inflation scenario by more than it falls in the high inflation scenario;

- Mortality increases with age, so the probability of surviving to higher ages falls rapidly.

These two effects combine to make the reduction in survival probability for the low inflation scenario greater than the increase in survival probability for the high inflation scenario. It follows that uncertainty in the rate of inflation may reduce the risk associated with a level annuity.

2.4 Stochastic Comparison

As the annual rate of inflation forms a sequence of random variables, it is difficult to quantify the risk of receiving inadequate income associated with a level annuity. Thus simulations are used to aid us in quantifying this risk.

In Sections 2.2 and 2.3 it was convenient to use a continuous-time model because most annuities bought in the U.K. provide a monthly income, making a continuous time model a reasonably good approximation of reality. In this section, however, we switch to a discrete time model for mathematical convenience. In addition, we assume that the annuity payments are made at the start of each year.

2.4.1 Inflation History of the United Kingdom

Parsons (1990) has examined the history of U.K. consumer price inflation since 1810 using the most representative index available in each era. From these data, he concludes that persistent positive inflation rates, as now exist in the U.K. and other industrialized economies, have only been observable since World War II. He therefore does not reject the possibility that inflation eventually may revert to its earlier pattern of behavior, in which the price level is as likely to fall as to rise in any
year and the average rate of inflation measured over long periods is small.

Another feature of the data is the existence of inflation shocks which appear to arise from the economic consequences of major historical events. Brief periods of high inflation occurred during all of the following crises:

- World War I;
- World War II;
- The rise in oil prices following the 1973 Arab/Israeli war; and
- The rise in oil prices at the start of Iran/Iraq war in 1980.

Currently, the rate of inflation is 3 percent per annum in the U.K., an historically low figure for the post-war era.

2.4.2 Scenarios for Future Inflation

For a pensioner retiring in current conditions there are three possible scenarios:

**Scenario 1:** The rate of inflation will continue to fall, and the economy will revert to long-term price stability;

**Scenario 2:** The economy will experience an inflation shock;

**Scenario 3:** Over the years, the rate of inflation will vary randomly within a few percentage points around a mean value close to the current rate.

The implications of the first two scenarios for the choice of annuity are clear. Under Scenario 1 fixed interest bonds would offer excellent real returns, and a level annuity would offer considerable extra value when compared with an index-linked annuity. Under Scenario 2 an index-linked annuity is the better choice to protect the real value of the pensioner's income. Under Scenario 3, however, the choice of annuity is unclear. Fortunately, this is the easiest scenario to model stochastically and is, perhaps, the most likely of the three scenarios to occur in the Britain.\(^{11}\)

\(^{11}\)In recent years a low and stable rate of inflation has been a stated objective of macroeconomic policy in Britain and other industrialized economies. The British government currently maintains a target range for inflation of 1 percent to 4 percent per annum.
2.4.3 Stochastic Model for Price Inflation

Stochastic models for the rate of inflation have been suggested by several authors including Wilkie (1986) and Clarkson (1991). Wilkie (1986) uses a first order autoregressive time series model\textsuperscript{12} for price inflation for \( t = 0, 1, 2, \ldots \), as follows:

\[
\ln(1 + b_{t+1}) = \mu_b + k_b[\ln(1 + b_t) - \mu_b] + \sigma_b Z_t \sqrt{1 - k_b^2}
\]

where \( b_t \) is the annual rate of inflation over year \( t \), i.e., over the interval \([t - 1, t)\) and the \( Z_t \)'s are a sequence of independent, identically distributed standard normal random variables (i.e., with mean zero and unit variance). In addition, \( \mu_b \) and \( \sigma_b \) are the mean and standard deviation of \( \ln(1 + b_t) \) respectively, and \( k_b \) is the correlation between \( \ln(1 + b_t) \) in adjacent years.

The Wilkie model assumes that inflation varies around a long-term mean, in accordance with Scenario 3. It has been criticized by Clarkson (1991) for failing to allow for inflation shocks and other nonlinear effects. The difficulty with incorporating shock terms is that past inflation shocks arose from disparate and unique historical events, so they offer little help in modeling the future shocks.

Huber (1995) has also pointed out that the data used by Wilkie to parameterize his model included inflation shocks, which is inappropriate given that the model assumes a constant mean and variance for the rate of inflation. As the assumption of stationarity is only valid for periods of moderate inflation, the parameters should be estimated from periods that exclude inflation shocks.

Wilkie's approach is used to model inflation under Scenario 3, but the parameters are estimated from U.K. inflation rates since 1982 (excluding the inflation shock of 1980/81). The parameters estimates are \( \mu_b = 0.047 \), \( \sigma_b = 0.019 \) and \( k_b = 0.58 \). We also require the expected inflation rate to be consistent with the bases used for pricing level and index-linked annuities. Thus, we ignore the estimate from past data and assume, instead, that \( \mu_b = \ln(1 + c) \). To summarize, the parameters used are

\[
\mu_b = \ln(1 + c), \quad \sigma_b = 0.019 \quad \text{and} \quad k_b = 0.58. \tag{9}
\]

It is possible that the variable \( c \), as calculated from equation (1), may not give a realistic estimate of future inflation if it is derived from actual annuity rates available in the market, which are influenced by factors such as mortality selection, expenses, and competition. If so, a different

\textsuperscript{12}For more on autoregressive time series, see, for example, Box and Jenkins (1976).
estimate could be used without affecting the method of comparison described in the next section.

2.4.4 Comparing Level and Index-Linked Annuities

For the deterministic comparison, we assume the pensioner buys a level annuity but only spends the income that would have been provided by an index-linked annuity, the difference being saved in (or withdrawn from) a sinking fund.

To simplify our simulations, we again assume the pensioner receives an income payable annually in advance. It follows that \( (AVS)_t \), the sinking fund per unit of retirement fund just before the annuity payment at time \( t \), must satisfy the following recurrence formula: \( (AVS)_0 = 0 \) and for \( t = 0, 1, 2, \ldots \)

\[
(\text{AVS})_{t+1} = \left( (\text{AVS})_t + \frac{1}{\delta} \right) (1 + b_{t+1}) (1 + r_{t+1})
\]

where \( r_t \) is the actual annual real interest earned on the sinking fund over the period \([t-1, t)\), \( b_t \) is the random actual annual rate of inflation over the period \([t-1, t)\), and

\[
\beta_t = \prod_{\tau=1}^{t} (1 + b_\tau)
\]

with \( \beta_0 = 1 \). The present value of the sinking fund, discounted using the actual interest rate earned on the sinking fund assets, is given by

\[
(PVS)_t = \frac{(\text{AVS})_t}{\beta_t R_t} \quad t = 0, 1, 2, \ldots
\]

where \( R_0 = 1 \) and

\[
R_t = \prod_{\tau=1}^{t} (1 + r_\tau) \quad t = 1, 2, \ldots
\]

At this point, the natural question to ask is what rate of interest should the sinking fund earn? The interest earned depends on the investment vehicles used. Given the importance of the sinking fund to the pensioner, it would be prudent to place sinking fund savings in a relatively secure, interest-bearing deposit account—preferably an account where the interest earned is linked to the rate of inflation, as the purpose of the sinking fund is to provide long-term protection against
inflation. In the U.K. appropriate instruments are Index-Linked National Savings Certificates, which provide a fixed real rate of interest over five year periods. The real interest rate offered for the seventh issue of these certificates in 1993 was 3 percent per annum, close to the average real yield on index-linked bonds.

We therefore assume that the sinking fund assets achieve a fixed real yield equivalent to that assumed in the pricing basis for index-linked annuities, i.e., that \( r_t = r \) for \( t = 1, 2, \ldots \).

### 2.4.5 Results of Simulations

For these simulations we assume that the sinking fund can go into debt when its assets have been exhausted. Let \( P_X(t, h) \) denote the probability that the present value of the sinking fund debt will exceed \( h \) at the end of \( t \) years, i.e.,

\[
P_X(t, h) = \Pr[(PVS)_t < -h].
\]

But this is also the probability that a sinking fund with initial assets at retirement equal to \( h \) would be exhausted after \( t \) years. Thus, \( P_X(t, h) \) is the probability that, even with additional savings of \( h \) at retirement, the pensioner will not be able to maintain an inflation-linked income for \( t \) years.

Next let \( P_X(h) \) be the probability that the sinking fund debt will exceed \( h \) at the end of the year of death. Assuming that the year of death does not depend on the rate of inflation, we can write:

\[
P_X(h) = \sum_{t=0}^{\infty} P_X(t+1, h) \cdot p_x q_{x+t}.
\]

To compute \( P_X(h) \), we proceed as follows:

**Step 1:** Perform 1000 simulations for realizations of \( (AVS)_t \) for \( t = 1, 2, 3, \ldots, 100 - x \) for selected retirement ages \( x \). These simulations are based on \( j = 7 \) percent and \( r = 3 \) percent and using the stochastic inflation model described in equations (8) and (9).

**Step 2:** For each \( x \) and \( t \), use the simulated values of to construct an empirical distribution function for \( (AVS)_t \).

**Step 3:** The empirical distribution functions are used to get a matrix of values of \( P_X(t, h) \) for \( h = 0.0, 0.1, 0.2, 0.3 \) and \( t = 1, 2, \ldots, 100 - x \).
Table 4

<table>
<thead>
<tr>
<th>$h$</th>
<th>$x$</th>
<th>0.00</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.38</td>
<td>0.17</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>0.33</td>
<td>0.15</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.31</td>
<td>0.14</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>0.29</td>
<td>0.14</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.24</td>
<td>0.11</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.22</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 shows the values of $P_x(h)$ estimated from our simulations. Notice that the probabilities when $h = 0$ are lower than those obtained for the deterministic comparison, as we might expect from our earlier results in Section 2.3. The other columns show the probability of exhausting the sinking fund when it contains assets at retirement equal to $h$ multiplied by the pensioner's retirement fund. An interesting feature of these simulations is that initial sinking fund assets equal to 20 percent of the pensioner's retirement fund would reduce the probability of exhaustion to 5 percent for all the retirement ages examined. Thus, for pensioners contemplating the purchase of a level annuity, additional liquid assets of roughly 20 percent of the retirement fund may insure reasonable protection against inflation in retirement, if one ignores the possibility of a severe inflation shock.

3 Income Withdrawal vs. Index-Linked Annuity

The income withdrawal option allows a pensioner to draw an income stream directly from the retirement fund instead of purchasing a whole life annuity. In the U.K. a pensioner who opts for income withdrawal can defer the purchase of a whole life annuity to age 75 at the latest.

The income withdrawal option is a means by which pensioners can avoid being locked into an asset offering an income linked to government bond yields at retirement, which many individuals may find unduly restrictive. Most pensioners opting for income withdrawal would choose to invest their fund in assets believed to offer higher returns (such as equities).
The risk premium, $\rho$, is defined as the difference between the expected real return on a diversified equity portfolio and the expected real yield on index-linked bonds. How large should the risk premium be? Wilkie (1994) considers this question and concludes that a figure of $\rho = 3$ percent per annum would be reasonable for the long-term risk premium on equities. Thus, if we assume a real yield of 3 percent in pricing index-linked annuities, it would be reasonable to assume an expected real return of 6 percent on our equity portfolio.

In this section we investigate the risks and benefits of drawing income from a fund invested entirely in equities by comparing the projected cash flows with those from an index-linked annuity. As in Section 2, we first adopt a deterministic approach using a continuous time model and then proceed to stochastic projections using a discrete time model.

### 3.1 Dividend Yield and Dividend Growth

An equity portfolio is expected to produce an increasing stream of dividend income. As equities are not redeemable, the expected return can be determined by evaluating the present value of the projected income stream over an infinite time horizon. We derive a formula for the real return from an equity portfolio in terms of the dividend yield and the rate of dividend growth, both of which are assumed to be constant. We first define the following terms for the equity portfolio:

- $V$ = Current market value of the equity portfolio;
- $d$ = Current rate of dividend income from the equity portfolio;
- $y = d/V$ = Dividend yield on the equity portfolio;
- $g$ = Real annual dividend growth; and
- $w$ = Real annual return on the equity portfolio.

Clearly, we must have $w > g$.

The return on the portfolio, $w$, is the rate of interest at which the present value of the projected dividend income from the portfolio is equal to its market value, hence:

$$V = d \int_0^\infty \left( \frac{1 + g}{1 + w} \right)^\tau d\tau$$

which implies
\[ 1 = y \int_0^\infty \left( \frac{1 + g}{1 + w} \right)^t dt \]

which further implies, after evaluating the integral, that

\[ \ln(1 + w) = y + \ln(1 + g) \]

Equation (15) implies that the interest earned on an equity portfolio can be split into two components: dividend yield and dividend growth. For the U.K. equity market as a whole, Thornton and Wilson (1992a) show that real dividend growth historically has averaged approximately 1 percent to 2 percent per annum. The balance between yield and growth, however, depends on the stocks selected; many equity funds have invested specifically to provide above average growth (hence lower yield) or above average yield (hence lower growth).

3.2 Deterministic Comparisons

3.2.1 Pensioners Who Live Off Dividends

A pensioner drawing income from an equity fund may wish to live off the dividends alone, to avoid selling stocks to meet income needs. For such a pensioner, a high yielding equity portfolio with zero real dividend growth\(^{13}\) may be preferable as the closest alternative to an index-linked annuity.

We consider an example where the real dividend growth is \(g = 0\), the risk premium on the equity portfolio is \(\rho = 3\) percent, \(\gamma = 3\) percent, and \(w = 6\) percent. Thus, it follows from equation (15) that the income from the equity portfolio per unit of retirement fund is \(y = \ln(1.06) \approx 5.8\) percent.

The comparable income yield from an index-linked annuity is \(1/\bar{a}_x\) which is shown in Table 5 for different retirement ages, using the male PA(90) life table. Table 5 shows that the dividend income from a high yielding equity portfolio is unlikely to match the income from an index-linked annuity at retirement ages above 50, the shortfall becoming greater as the age of retirement increases.

After retirement the real value of the fund will not change, given our assumption of a flat dividend yield and zero real dividend growth. It follows that when the pensioner later buys an annuity, the same fund (in

\(^{13}\)An example of such a portfolio is the M&G Equity Income Fund, which in providing an above average income yield for its investors has achieved income growth roughly line with price inflation since its formation in 1972. M&G is one of the leading unit trusts (mutual funds) in the U.K.)
real terms) will be available to purchase a cheaper annuity (because the pensioner is older). If the annuity is bought \( m \) years after retirement, the income (compared with buying an annuity at retirement) increases by the proportion:

\[
\Delta_m = \frac{\bar{a}_{\bar{r}}_x}{\bar{a}_{x+m}} - 1. \tag{16}
\]

Table 6 shows the percentage increase in income obtained by deferring the purchase of an annuity, for selected values of \( x \) and \( m \).

### Table 5

**Index-Linked Annuity Income Yield**

<table>
<thead>
<tr>
<th>( x )</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield</td>
<td>5.2%</td>
<td>5.8%</td>
<td>6.6%</td>
<td>7.6%</td>
<td>8.9%</td>
<td>10.8%</td>
<td>13.4%</td>
</tr>
</tbody>
</table>

As expected from equation (16), a greater increase in income is achieved for a pensioner who retires later and defers the annuity purchase longer. The sacrifice of income before the annuity purchase, however, also increases with retirement age (Table 5). In the next section we use a method of comparison to determine when a higher overall income can be obtained by drawing income from an equity fund.

#### 3.2.2 Pensioner Who Matches Annuity Income

We now assume the pensioner draws an income from the fund that matches the income from an index-linked annuity, selling assets (or reinvesting surplus dividends) as required. If the pensioner can obtain a
higher income after the purchase of an annuity, the income withdrawal option would be advantageous.

Using currency units that are linked to price inflation, we now define, for \( t \geq 0 \), the following:

\[
 f_t = \text{Real value of the fund at time } t, \text{ per unit of the initial fund at time } 0.
\]

We assume the pensioner retires at \( t = 0 \), hence \( f_0 = 1 \). If assets are continually sold to maintain the same income as from an index-linked annuity, \( f_t \) must satisfy the following differential equation:

\[
 -\frac{df_t}{dt} + yf_t = \frac{1}{\bar{a}_x^y}.
\]

(17)

Equation (17) yields the solution

\[
 f_t = e^{yt} - \frac{s^y_{|t|}}{\bar{a}_x^y}
\]

(18)

where \( s^y_{|t|} = (e^{yt} - 1)/y \). The proportionate change in income on purchasing an annuity at age \( x + m \) is given by:

\[
 \Delta_m = f_m \frac{\bar{a}_x^y}{\bar{a}_{x+m}^y} - 1.
\]

(19)

Table 7 shows the increase in income on buying an annuity for different \( x \) and \( m \). Table 7 suggests there is a critical retirement age above

<table>
<thead>
<tr>
<th>( x )</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>15.8%</td>
<td>38.1%</td>
<td>69.3%</td>
<td>115.4%</td>
<td>185.9%</td>
<td>297.1%</td>
</tr>
<tr>
<td>50</td>
<td>14.0%</td>
<td>32.2%</td>
<td>56.8%</td>
<td>91.0%</td>
<td>139.6%</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>10.5%</td>
<td>22.6%</td>
<td>35.7%</td>
<td>47.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>6.0%</td>
<td>8.9%</td>
<td>3.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>-0.8%</td>
<td>-12.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>-11.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
which the pensioner suffers from taking the income withdrawal option. For pensioners who defer the purchase of the annuity until age 75, the projected increase in income is only 3 percent for a retirement age of 60 and falls to a reduction of 13 percent for a retirement age of 65. Thus, drawing income from an equity fund cannot be expected to provide a greater overall income for retirement ages much above 60 using the male PA(90) life table to price annuities.

It does not follow, however, that pensioners retiring at younger ages should always opt for income withdrawal because the investment risk involved may be unacceptable. In the next section we attempt to quantify this risk through simulation, using a stochastic model for the return obtained from the equity portfolio.

3.3 Stochastic Comparison

Deterministic projections, as used in Section 3.3, can be used to find when a fund invested in equities is likely to provide more income than a whole life annuity and the expected amount of this extra income. In this section we use stochastic projections to estimate:

- The probability that the pensioner is unable to match the income of an index-linked whole life annuity as a result of poor investment experience over the period in which income is drawn from the equity fund; and
- The amount of additional savings at retirement that would give reasonable assurance of maintaining an inflation-linked income in spite of poor investment experience.

3.3.1 Formulae for Projections

We require a stochastic model that will enable simulation of the market value of our equity portfolio and its dividend income. As in Section 2.4, we switch to a discrete time framework in which a stochastic approach is more easily accommodated.

For \( t = 1, 2, \ldots \), let

\[
V_t = \text{Real market value of equity portfolio at time } t;
\]

\[
d_t = \text{Real dividend income from equity portfolio paid at time } t;
\]

\[
y_t = \frac{d_t}{V_t} = \text{Dividend yield on the equity portfolio at time } t;
\]

\[
i_t = \text{Real actuarial return on fund over } [t - 1, t);
\]
\[ c_t = \text{Real growth in market value of assets over } [t-1, t); \text{ and} \]
\[ g_t = \text{Real annual dividend growth over } [t-1, t). \]

As \( d_t = d_{t-1}(1 + g_t) \) and \( 1 + c_t = V_t/V_{t-1} \), we get
\[ c_t = (1 + g_t) \frac{Y_{t-1}}{Y_t} - 1. \] (20)

Assuming that the pensioner withdraws or reinvests assets as needed to match the income from an index-linked annuity and that the sale or purchase of assets does not alter the composition of the portfolio (i.e., the relative weighting given to each individual stock does not change), leads to the following recurrence formula for \( f_t \):
\[ f_{t+1} = \left( (1 + y_t)f_t - \frac{1}{\delta x} \right) (1 + c_{t+1}). \] (21)

Following Thornton and Wilson (1992b) we define the real actuarial return on the fund as:
\[ i_t = (1 + y_t)(1 + g_t) - 1. \] (22)

Using equations (20) and (22), equation (21) can be rewritten as:
\[ f_{t+1} = \left( (1 + y_t)f_t - \frac{1}{\delta x} \right) \frac{Y_t(1 + i_{t+1})}{Y_{t+1}(1 + y_t)}. \] (23)

### 3.3.2 Stochastic Model

Our stochastic model for the equity portfolio consists of the following two components:

- The factors \( 1 + i_t \) form a sequence of independent, identically distributed log-normal random variables, i.e., for \( t = 1, 2, \ldots \):
  \[ \ln(1 + i_t) \sim N(\mu_t, \sigma_t^2); \text{ and} \]

- The sequence \( \ln(y_t) \) follows a first order autoregressive process with a normal residual for \( t = 0, 1, 2, \ldots \), as follows:
  \[ \ln(y_{t+1}) = \mu_y + k_y \left( \ln(y_t) - \mu_y \right) + \sigma_y Z_t \sqrt{1 - k_y^2} \] (24)
  where \( \mu_y, \sigma_y \) and \( k_y \) are parametric constants representing, respectively, the mean of \( \ln(y_t) \), its standard deviation, and the correlation between the \( \ln(y_t) \)'s in adjacent years.
The first component of the model focuses on the actuarial return rather than the return on market value, as historic data for the U.K. equity market show that actuarial returns have been much less variable. Thus, a model based on actuarial returns is likely to be a better description of the behavior of the U.K. equity market. The second component of the model is similar to the approach used by Wilkie in assuming that the dividend yield on U.K. equities tends to revert to a long-term average. This implies that the equity market tends to correct itself when stock prices are overvalued or undervalued relative to some par dividend yield, an assumption well supported by historic data for the U.K. equity market.

As in Section 3.3, we assume the pensioner invests in a portfolio of high yielding stocks from which the expected real dividend growth is zero. If dividends are payable annually in advance, the real return on the equity portfolio, $w$, satisfies the equation:

$$V_0 = d_0 \sum_{t=0}^{\infty} (1 + w)^{-t} = d_0 \frac{(1 + w)}{w},$$

which implies that $\gamma_0 = w/(1 + w)$. We also must assume that the long-term par dividend yield is consistent with $w$, i.e.,

$$\mu_y = \ln(\gamma_0) = \ln\left[\frac{w}{(1 + w)}\right].$$

Equation (22) implies that when real dividend growth is zero, the actuarial return over any year is equal to the dividend yield at the start of the year. We therefore use the following estimate for $\mu_i$:

$$\mu_i = \ln(1 + \gamma_0) = \ln\left[1 + \frac{w}{(1 + w)}\right].$$

Values for the other parameters used in our stochastic model are estimated from representative U.K. equity indices from 1919 to 1995\textsuperscript{14} as follows:

$$\sigma_i = 0.0675, \quad \sigma_d = 0.24, \quad \text{and} \quad k_d = 0.50$$

which, respectively, are the standard deviation of the logarithm of the actuarial return, the standard deviation of the logarithm of the dividend yield at each year-end, and the correlation between the logarithm of dividend yields in adjacent years.

\textsuperscript{14}The index used is the BZW equity index, a representative U.K. stock price index compiled by the investment bank, Barclays de Zoete Wedd. The index provides data on U.K. stock prices and yields from 1919 to the present.
3.3.3 Present Value of Surplus

We use the model described in Section 3.4.2 above to simulate values for the fund, \( f_t \), at different durations from retirement. For each projection, we are interested in comparing the market value of the fund at any chosen duration with the money required to maintain an unchanged income after the purchase of an index-linked annuity.

In carrying out these simulations there are two complications:

- The pricing basis for index-linked annuities may change over time;

- The pensioner can time the purchase of the annuity to exploit favorable changes in investment yields.

Strictly, we also require a stochastic model for the real yield on index-linked bonds to allow for random fluctuations in the pricing basis. Given that index-linked yields have been more stable than equity dividend yields and that most of the uncertainty is believed to arise from variability in equity returns, however, it may be adequate for our purpose to assume that the pricing basis does not alter between retirement and the purchase of the annuity.

We also ignore the second complication, for it assumes that pensioners can judge when equities are overpriced relative to index-linked bonds, something that even experienced fund managers may find difficult. As in Section 3.2, we perform projections assuming that the pensioner defers the purchase of an annuity for a fixed period.

On purchasing an annuity \( m \) years after retirement the surplus assets, \( u_m \), are given by:

\[
 u_m = f_m - \frac{\bar{a}_x^r + m}{\bar{a}_x^r}.
\]  

(27)

If \( u_m = 0 \) the fund would be just sufficient to purchase an index-linked annuity providing the same real income. We assume that \( u_m \) can be negative as well as positive, which is implied from the use of our stochastic model.

Let \((PVU)_x(m)\) be the present value at retirement of the projected surplus or deficit using the same real interest used to price the index-linked annuities, i.e.,

\[
(PVU)_x(m) = \frac{u_m}{(1 + r)^m}.
\]  

(28)
3.3.4 Results of Simulations

Next we use simulations to estimate

\[ m \xi_x(-h) = \Pr[(PVU)_x(m) < -h \mid \text{pensioner survives to age } x + m] \]

for \( h = 0, 0.1, 0.2, 0.3 \). Note, for example, that \( \Pr[(PVU)_x(m) < 0] \) gives the probability that a pensioner who retires at age \( x \) will not be able to maintain the same real income after using the remaining fund to purchase an index-linked annuity \( m \) years after retirement.

One thousand simulations are performed for selected values of \( x \) and \( m \) using the stochastic model described in Section 3.4.2 and assuming: \( r = 3 \) percent and \( w = 6 \) percent. The initial dividend yield for each simulation and its long-term average value are as given in equation (25).

Table 8 is consistent with Table 6 in showing that the income withdrawal option becomes more risky as the age of retirement increases. The risk involved is significant even at a retirement age as low as 45. Tables 9, 10, and 11 show the estimated probabilities for \( h = 0.1, 0.2, \) and 0.3, respectively.

<table>
<thead>
<tr>
<th>( x )</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.31</td>
<td>0.22</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>50</td>
<td>0.33</td>
<td>0.26</td>
<td>0.23</td>
<td>0.20</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>0.40</td>
<td>0.33</td>
<td>0.34</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.44</td>
<td>0.43</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>0.49</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using the same reasoning as Section 2.4.5, Tables 9, 10, and 11 give probabilities of not being able to maintain an inflation-linked income for a pensioner who had additional savings at retirement equal to \( h \) times the retirement fund and invested these savings in assets giving a guaranteed real return of 3 percent per annum.

Tables 8 through 11 show that each increment of 10 percent in the additional savings held at retirement significantly reduces the risk for any combination of \( x \) and \( m \). For initial savings of 30 percent of the retirement fund, the risk is small for the younger retirement ages.
Table 9
Simulation Estimates of $m\xi_x(-0.1)$

<table>
<thead>
<tr>
<th>$x$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.16</td>
<td>0.11</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>50</td>
<td>0.18</td>
<td>0.13</td>
<td>0.13</td>
<td>0.09</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>0.23</td>
<td>0.17</td>
<td>0.19</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.28</td>
<td>0.27</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>0.32</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10
Simulation Estimates of $m\xi_x(-0.2)$

<table>
<thead>
<tr>
<th>$x$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>50</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.12</td>
<td>0.13</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>0.17</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Summary and Conclusions

As expected, the three important factors that affect how a pensioner's retirement fund should be invested to provide an index-linked income stream for life are expected future price inflation, the pensioner's expected remaining life span, and the additional savings held by the pensioner. Apart from purchasing and index-linked annuity, we assume that the pensioner can either purchase a level whole life annuity or select income withdrawal.

4.1 Level Whole Life Annuities

The existence of price inflation has meant that the traditional whole-life annuity, providing a level monetary income, is no longer a risk-free option. Thus, we have adopted the newer index-linked annuity as the benchmark against which other options should be measured.
Table 11  
**Simulation Estimates of** $m \xi_x(-0.3)$  

<table>
<thead>
<tr>
<th>$x$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>50</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.03</td>
<td>0.04</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>0.06</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most U.K. pensioners still opt for a level annuity, however, and this is not an irrational preference. We show that the odds are in favor of obtaining a higher aggregate income from a level annuity, especially at older retirement ages. We also show that the existence of price inflation poses a significant longevity risk to the recipient of a level annuity, irrespective of whether the rate of inflation is constant or random. Thus, the purchase of a level annuity is perhaps only advisable for individuals with additional savings; we estimate that further assets of at least 20 percent of the retirement fund are necessary to insure against the risk of being unable to maintain an index-linked income stream until death.

4.2 Income Withdrawal

Our analysis of the income withdrawal option can be summarized as follows:

- The expected overall income from the equity fund is greater for retirement ages below a critical threshold at which the extra return from investing in equities is balanced by the extra cost of capital protection on death. Using the male PA(90) life table, this critical age is somewhere between 60 and 65;

- The expected extra income from the equity fund reduces as age of retirement increases, becoming more negative as the retirement age increases beyond the critical age;

- Even at retirement ages as young as 50, there remains a significant risk of underperforming relative to an index-linked annuity as a result of poor equity returns;
• Additional savings invested in assets providing a guaranteed real return can significantly reduce the risk of not being able to match the income from an index-linked annuity—savings of 30 percent of the retirement fund would reduce the risk to below 5 percent for retirement ages at which a higher overall income was expected.

Thus the greater the life expectancy at retirement, the greater the advantages of drawing income from an equity fund compared with a whole life annuity providing an income linked to bond yields. This accords with the actuarial viewpoint that equities are suitable assets for matching longer-term liabilities.

There is still a significant risk, however, that an equity fund may not be able to match the income from an index-linked annuity, even at young retirement ages. Again, the pensioner should have additional savings at retirement, with the minimum savings required varying between 20 percent and 30 percent of the retirement fund.

As a pensioner's actual life span is uncertain, the longevity risk associated with drawing income from a fund eventually will dominate other considerations. Thus, at some age the purchase of a whole life annuity becomes necessary. U.K. legislation does not allow the purchase of an annuity to be deferred beyond age 75; we have shown that a male pensioner may be unwise to defer the purchase beyond age 65.

References


Pension Funding by Normal Costs or Amortization of Unfunded Liabilities

Keith P. Sharp*

Abstract†

We discuss the extent of the actuary's freedom in choosing the funding method for defined benefit pension plans. In particular, we look at funding through a combination of normal costs, amortization of an unfunded liabilities, and fund of assets. The IRS constraint on "reasonable funding methods" is considered, with particular mention of the aggregate entry age normal method. In addition, an algebraic development is performed of year-to-year changes in the status of a plan's funding.

Key words and phrases: reasonable funding

1 Introduction

There are many methods used by actuaries to evaluate the funding of defined benefit pension plans. The choice of funding method is influenced by several factors, including:

- The plan's benefit design; in particular, whether the pension benefit is related to final salary;

- The plan sponsor's objectives;

*Keith Sharp, F.S.A., Ph.D., is an associate professor at the University of Waterloo, Canada. He has worked as an actuary for pension consulting firms in Canada and Australia and for an insurance company in Britain. His papers have appeared in various journals, including Transactions of the Society of Actuaries, Insurance: Mathematics and Economics and Journal of Risk and Insurance.

Dr. Sharp's address is: Department of Statistics & Actuarial Science, University of Waterloo, Waterloo ON N2L 3G1, CANADA. Internet address: 73231.16@compuserve.com

†The author acknowledges the valuable contributions of the anonymous referees. All errors remain his responsibility.

257
• The requirements under the appropriate regulatory environment; and

• The traditions of the geographic area and of the actuary's firm.

This paper explores the extent of the actuary's freedom in devising methods for funding benefits and in adjusting contributions to take account of gains and losses and plan improvements. A particular constraint considered is the IRS requirement for "reasonable funding methods" to be used. Details of the mathematical characteristics of such reasonable funding methods are given in Appendices A through E.

Appendix A considers the definition of accrual (actuarial) liability for benefit allocation methods and shows the equivalence of the present value of accrued benefits and the $AL = PVFB - PVFNC$ definitions. Appendix B considers the frozen initial liability and aggregate methods, with their definitions of unfunded liability. Appendix C gives a more thorough confirmation that the benefit allocation methods adhere to the zero-gain criterion. Appendix D indicates that the individual level cost methods, too, satisfy the zero-gain criterion. Appendix E discusses the non-individual methods: in other words those in which the numerator and the denominator defining normal cost are separately summed over plan members. Thus, the frozen initial liability and aggregate methods are considered in Appendix E. Finally, Appendix F contains a numerical example.

2 Benefit Value as a Sum of Components

Fundamental to the actuarial valuation of a pension plan is that the actuary must ensure that the present value of projected future benefits at any time $t (PVFB_t)$ be balanced by the sum of the plan's assets of various types. Available assets (tangible and intangible) for a valuation at time $t$ are:

• $F_t$, the fund of tangible invested assets at actuarial value, possibly a smoothed market value at time $t$;

• $PVFNC_t$, the present value at time $t$ of future normal costs for plan members at the valuation at $t$, based on their normal costs ($NC_t$) calculated at that valuation at time $t$;

• $UAL_t$, the unfunded actuarial liability at time $t$. It is based on the initial unfunded actuarial liability $UAL_0$, which is amortized by level dollar annual payments $UAL_0/\tilde{a}_{\bar{m}}$. As a result, the $UAL$ can
be regarded as the intangible asset of the present value of future amortization payments.

It follows that the equation of value that must be satisfied by any method for the entire pension plan is

$$PVFB_t = PVFNC_t + UAL_t + F_t.$$  \hspace{1cm} (1)

Substituting $UAL_t = AL_t - F_t$ gives the usual expression for the plan’s accrued liability at $t$:

$$AL_t = PVFB_t - PVFNC_t.$$  

Thus, for a funding arrangement to be satisfactory, it is necessary but not sufficient for equation (1) to hold. This is considered in more detail in Section 3.

Because we need notation to allow for the various versions of quantities at any given time, the notation described in Table 1 is used for quantities at time $t$. Note, all quantities refer to the sum over plan members.

Some actuaries may prefer that calculations be done based on calculating the cost of plan modifications on the revised assumptions rather than on the previous assumptions. The results of this paper can be readily modified by regarding $M$ as denoting modified assumptions and $R$ as denoting a revised plan.

Consider column (2) of Table 2. The time $t - 1$ plan normal cost $NC_{t-1}^R$ is based on the time $t - 1$ revised assumptions and on the time $t - 1$ plan document with any amending modifications. Making the assumption that normal costs are payable at the beginning of the year, we have for the whole plan

$$PVFNC_{t-1}^R \times (1 + i) = NC_{t-1}^R \times (1 + i) + PVFNC_t^E$$  \hspace{1cm} (2)

where $PVFNC_t^E$ is the notation for the present value at time $t$ of the normal costs expected at time $t - 1$. The quantities denoted by $E$ at time $t$ are the same as those denoted by $R$ at time $t - 1$. The validity of equation (2) is apparent for any predefined series of payments, including changes in the membership.

Experience may differ from assumed in various areas including the number of terminations and the amount of salary increases. Taking this into account, the time $t$ present value of future normal costs with gains or losses is given by

$$PVFNC_t^G = PVFNC_t^E + NC_t$$  \hspace{1cm} (3)
Table 1
Summary of Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$EOV_{t-1}$</td>
<td>Equation of value at $t - 1$.</td>
</tr>
<tr>
<td>$inv^G$</td>
<td>Investment component of gain;</td>
</tr>
<tr>
<td>$\Delta^M$</td>
<td>Change resulting from modification to the plan document;</td>
</tr>
<tr>
<td>$\Delta^R$</td>
<td>Change resulting from revisions to assumptions;</td>
</tr>
<tr>
<td>$F_t$</td>
<td>Invested assets; there may be cash contributed to fund modifications or revisions to assumptions resulting in a new fund $F^R$.</td>
</tr>
</tbody>
</table>

Superscript Notation

<table>
<thead>
<tr>
<th>Superscript</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>Expected outcome if time $t - 1$ assumptions are realized;</td>
</tr>
<tr>
<td>$G$</td>
<td>Actual outcome at time $t$ with inclusion of gains (or losses) since the previous valuation, assumed to be at $t - 1$;</td>
</tr>
<tr>
<td>$M$</td>
<td>Includes modifications effective at time $t$ to the plan document;</td>
</tr>
<tr>
<td>$R$</td>
<td>Includes revisions effective at time $t$ to the actuarial assumptions; and</td>
</tr>
<tr>
<td>$A$</td>
<td>Any one of the above $E$, $G$, $M$, or $R$;</td>
</tr>
</tbody>
</table>

where $NC G_t$ is the portion of gain related to changes in the payroll on which the normal cost is calculated. Plan modifications at time $t$ may cause another change $\Delta^M PVFNC_t$ to give the quantity $PVFNC^M_t$ including modifications:

$$PVFNC^M_t = PVFNC^G_t + \Delta^M PVFNC_t.$$ (4)

Similarly, including the effect of assumptions revised as of time $t$, we have

$$PVFNC^R_t = PVFNC^M_t + \Delta^R PVFNC_t.$$ (5)

Thus, we have confirmed column (2) of Table 2.

Column (3) of Table 2 indicates the development of the unfunded accrued (actuarial) liability over time. Changes in the unfunded may result from:
Table 2  
Development of Asset Components of $PVFB$  
For the Entire Pension Plan

<table>
<thead>
<tr>
<th>Description</th>
<th>Intangible</th>
<th>Intangible</th>
<th>Invested</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$EOV_{t-1}$</td>
<td>$PVFNC^R_{t-1}$</td>
<td>$UAL^R_{t-1}$</td>
<td>$F^R_{t-1}$</td>
</tr>
<tr>
<td>$(1 + i)EOV_{t-1}$</td>
<td>$PVFNC^R_{t-1}(1 + i)$</td>
<td>$UAL^R_{t-1}(1 + i)$</td>
<td>$F^R_{t-1}(1 + i)$</td>
</tr>
<tr>
<td>Contribution:</td>
<td>$-NC^R_{t-1}(1 + i)$</td>
<td>$- (iC_{t-1} - NC^R_{t-1}(1 + i))$</td>
<td>$+iC_{t-1}$</td>
</tr>
<tr>
<td>Benefits:</td>
<td>$+NC \cdot G_t$</td>
<td>$NJ^{tot}G_t$</td>
<td>$+inv \cdot G_t$</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>$PVFNC^E_t$</td>
<td>$UAL^E_t$</td>
<td>$F^E_t$</td>
</tr>
<tr>
<td>Gain</td>
<td>$+NC \cdot G_t$</td>
<td>$NJ^{tot}G_t$</td>
<td>$+inv \cdot G_t$</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>$PVFNC^G_t$</td>
<td>$UAL^G_t$</td>
<td>$F^G_t$</td>
</tr>
<tr>
<td>Modifications:</td>
<td>$+\Delta^M PVFNC_t$</td>
<td>$+\Delta^M UAL_t$</td>
<td>$+\Delta^M F_t$</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>$PVFNC^M_t$</td>
<td>$UAL^M_t$</td>
<td>$F^M_t$</td>
</tr>
<tr>
<td>Revisions:</td>
<td>$+\Delta^R PVFNC_t$</td>
<td>$+\Delta^R UAL_t$</td>
<td>$+\Delta^R F_t$</td>
</tr>
<tr>
<td>Total:</td>
<td>$PVFNC^R_t$</td>
<td>$UAL^R_t$</td>
<td>$F^R_t$</td>
</tr>
</tbody>
</table>

$EOV_{t-1} = \text{Equation of value at } t - 1; \text{ Column (2) = Equations (2) to (5); Column (3) = Equation (11); Column (4) = Equation (12).}$

- Experience gains or losses;
- Plan amendments;
- Plan inception (a special case of a plan amendment); or
- Changes in assumptions

as of $t - 1$ or an earlier date. The unfunded liability may be under amortization and be regarded as an intangible asset equal to the present value of the scheduled amortization payments.

Following the notation used by Aitken (1994, p. 150), $iC_{t-1}$ denotes the actual contributions for the year $[t - 1, t)$ accumulated to $t$ at the assumed rate $i$. Also, $iC_{t-1} - NC^R_{t-1}(1 + i)$ is regarded as a supplemental cost (McGill and Grubbs, 1989), which reduces the unfunded liability:

$$UAL^E_t = UAL^R_{t-1}(1 + i) - (iC_{t-1} - NC^R_{t-1}(1 + i)).$$ (6)
The unfunded liability at \( t \) may be the sum of several previous unfunded liabilities that are being amortized over different periods. The unfunded liability may consist only of the \( n \) year level dollar amortization of an unfunded liability detected at time \( T \). Then we may have level contributions

\[
i C_t = \left( \frac{UALT}{\bar{a}_{n|}} + NC_t \right)(1 + i)
\]

and the amortization of equation (6) proceeds as

\[
UALT_t = \left( UALT_{t-1} - \frac{UALT}{\bar{a}_{n|}} \right)(1 + i).
\]

This special case thus gives the familiar formula for a level dollar amortization of the component \( UALT \) of the unfunded liability:

\[
UALT_t = UALT_T \frac{\bar{a}_{n-(t+T)}}{\bar{a}_{n|}}.
\]

Let us now consider the gain \( \text{tot} G \).\(^1\) Thus the end of year unfunded is \( UALT_{t-1} (1 + i) \) reduced by the degree \( \left( i C_{t-1} - NC_{t-1} (1 + i) \right) \) to which actual contributions exceed that normal cost.

The end of year unfunded is further reduced by any total (i.e., investment, decrements etc.) gain \( \text{tot} G_t \) to give

\[
UALT_t = \left( UALT_{t-1} + NC_{t-1} \right)(1 + i) - i C_{t-1} - \text{tot} G_t.
\]

Equation (10) corresponds to the top four cells of column (3) of Table 2. It is often expressed as a formula for the gain, taken to the left side. It can be approached from various directions. (See e.g., Aitken, 1994, p. 157, and Anderson, 1992, p. 13.) The expression \( UALT_t \) indicates the time \( t \) balance after gains or losses but before any time \( t \) amendments or changes in assumptions. Such changes add amounts \( \Delta^M UALT \) and \( \Delta^R UALT \) respectively to give column (3) of Table 2:

\[
UALT_t = UALT_{t-1} (1 + i) - (i C_{t-1} - NC_{t-1} (1 + i)) - \text{tot} G_t + \Delta^M UALT + \Delta^R UALT.
\]

\(^1\) The gain \( \text{tot} G \) can be regarded as the amount by which the actual end of year unfunded is less than the expected (if all assumptions were realized) end of year unfunded. In addition, the normal cost can be defined as the contribution that would result in the unfunded normally growing with interest. Here \( \text{normally} \) is interpreted as all assumptions being realized.
Table 3
Development of Components of PVFB

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( EOV_{t-1} )</td>
<td>( PVFB_{t-1}^{R} )</td>
<td>( AL_{t-1}^{R} )</td>
</tr>
<tr>
<td>((1 + i) EOV_{t-1})</td>
<td>( PVFB_{t-1}^{R} (1 + i) )</td>
<td>( AL_{t-1}^{R} (1 + i) )</td>
</tr>
<tr>
<td>Contribution:</td>
<td>+0</td>
<td>+( NC_{t-1}^{R} (1 + i) )</td>
</tr>
<tr>
<td>Benefits:</td>
<td>(-iB_{t-1})</td>
<td>(-iB_{t-1})</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>( PVFB_{t}^{E} )</td>
<td>( AL_{t}^{E} )</td>
</tr>
<tr>
<td>Gain</td>
<td>(- (\text{tot } G_{t} - \text{inv } G_{t} - NC_{t} G_{t}) )</td>
<td>(- (\text{tot } G_{t} - \text{inv } G_{t}) )</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>( PVFB_{t}^{E} )</td>
<td>( AL_{t}^{E} )</td>
</tr>
<tr>
<td>Modifications:</td>
<td>+( \Delta^{M} PVFB_{t} )</td>
<td>+( \Delta^{M} AL_{t} )</td>
</tr>
<tr>
<td>Sub-Total:</td>
<td>( PVFB_{t}^{M} )</td>
<td>( AL_{t}^{M} )</td>
</tr>
<tr>
<td>Revisions:</td>
<td>+( \Delta^{R} PVFB_{t} )</td>
<td>+( \Delta^{R} AL_{t} )</td>
</tr>
<tr>
<td>Total:</td>
<td>( PVFB_{t}^{R} )</td>
<td>( AL_{t}^{R} )</td>
</tr>
</tbody>
</table>

\( EOV_{t-1} \) = Equation of value at \( t - 1 \); Column (2) = Columns (3) + (4) of Table 1; Column (3) = Columns (2) + (3) + (4) of Table 1.

Column (4) of Table 2 indicates the fund being increased by contributions and reduced by benefits (and expenses if paid by the fund). Interest to the end of the year is calculated at the annual rate. The necessary correction for reality is the investment gain \( \text{inv } G_{t} \). This investment gain is identical to the excess of interest earned on a savings account over the amount that would have been earned at some assumed rate \( i \). One can allow for the possibility of a lump sum contribution of amount \( \Delta^{M} F_{t} + \Delta^{R} F_{t} \) to give column (4) of Table 2:

\[
F_{t}^{R} = F_{t-1}^{R} (1 + i) + iC_{t-1} - iB_{t-1} + \text{inv } G_{t} + \Delta^{M} F_{t} + \Delta^{R} F_{t}.
\] (12)

Column (2) of Table 3 shows the breakdown of the change from \( PVFB_{t-1}^{R} \) to \( PVFB_{t}^{R} \) and, consistently, is the total of columns (2), (3), and (4) of Table 2. Column (3) of Table 3 gives the development of the accrued liability \( AL \) and equals column (2) of Table 3 less column (2) of Table 2; \( AL = PVFB - PVFNC \).
3 Desirable Characteristics of a Funding Method

Legal requirements of the applicable jurisdiction must be satisfied together with the code of conduct requirements of the actuary’s professional body. Other considerations will also come into play. Among the matters to consider in choosing a funding method are:

- Rate of funding of accruing benefits;
- Speed at which the cost of plan improvements, including plan inception, is amortized;
- Degree to which cost to the employer is level and predictable, perhaps as a percentage of payroll; and
- Degree to which a surplus or unfunded liability is produced.

In recent years there has been much focus on the question of pension plan surplus. The employer may be required to make up, for example, any shortfall of assets on plan termination. But in some jurisdictions (e.g., Ontario, Canada) the employer may have difficulty in recovering any surplus. In view of this one-way bet, some affected employers may tend to favor low rates of contribution even though this reduces the security of benefits.

The speed of funding may have significant consequences. Consider an extreme example that may not be allowed under IRS regulations. Membership includes a highly compensated individual age 64 at valuation at \( t \). Pensions are paid by annuity purchase rather than monthly withdrawals from the fund. An assumed age 65 retirement could, under equation (1), be balanced by normal costs payable over an extended future period. But on the retirement there may be insufficient invested assets to purchase the large required annuity. Thus, in reality, the intangible assets \( PVFNC_t^A \) and \( UAL_t^A \) cannot always substitute for the invested asset \( F_t^A \). Attention must be paid to the incidence and not only to the present value of the normal cost and amortization payment streams.

4 IRS Reasonable Funding Method

According to §1.412(c)(3) - 1(c)(2) of regulations under the Internal Revenue Code, under a reasonable funding method no experience gains or losses are produced if each actuarial assumption is exactly realized. Below we consider which classes of methods satisfy this zero-gain criterion.
Let us examine the gain for methods that satisfy
\[ PVFNC_t^R + UAL_t^R + F_t^R = PVFB_t^R. \] (13)
We assume all plan assumptions are realized; so, for example, the equation
\[ \left[ PVFNC_{t-1}^R - NC_{t-1}^R \right](1 + i) = PVFNC_t^G, \]
which corresponds to column (2) of Table 2, is satisfied.

We start from the standard definition of total gain (Aitken, 1994, p. 50, and Anderson, 1992, p. 20)
\[ \begin{align*}
  \text{tot} G_t &= (UAL_{t-1}^R + NC_{t-1}^R)(1 + i) - UAL_t^G - iC_{t-1} \nonumber \\
  &= (PVFB_t^R - PVFNC_{t-1}^R - F_{t-1}^R + NC_{t-1}^R)(1 + i) \nonumber \\
  &\quad - (PVFB_t^G - PVFNC_t^G - F_t^G) - iC_{t-1} \nonumber \\
  &= iB_t - iC_{t-1} - F_{t-1}^R(1 + i) + F_t^G \nonumber \\
  &= 0. 
\end{align*} \]

In the above expressions for the total gain, we have assumed that there are no plan modifications or revisions to assumptions, and we used column (2) of Table 2 and (2) of Table 3. Thus, we have shown that methods satisfying equation (13) satisfy the zero-gain criterion.

From column (4) of Table 2 one finds that a zero investment gain results if the assumed rate of interest is realized. It can therefore be useful to concentrate on the non-investment portion of gain. From column (3) of Table 3 we can find an expression for the non-investment (or liability) portion of the total gain:
\[ \begin{align*}
  \text{tot} G_t - \text{inv} G_t &= AL_{t-1}^R(1 + i) + NC_{t-1}^R(1 + i) - iB_{t-1} - AL_t^G. \quad (14) 
\end{align*} \]
Equation (14) can be used to examine whether a method satisfies the zero-gain criterion.

5 Aggregate Entry Age Normal

The plan normal cost under the aggregate entry age normal method is defined (Aitken, 1994, p. 131 and Daskais, 1982) as
\[ NC_t^R = n_t \times \frac{\sum M_t PV_{e_j} FB_t^R}{\sum M_t d_{e_j}^R e_j \gamma - e_j}. \] (15)
where \( n_t \) is the number of active members at time \( t \), \( e_j \) is the plan entry date of employee \( j \) and \( M_t \) is the set of active members at time \( t \). \( PV_{ej} FB^R \) is the present value at entry age of the benefits of employee \( j \), including the effect of any plan revisions. It is necessary that equation (13) be obeyed so we have an unfunded liability given for the plan by

\[
UAL_t^R = PVFB_t^R - PVFNc_t^R - F_t^R.
\]  

Calculation of \( PVFNc_t^R \) is complicated in view of the equation (15) definition of plan normal cost and future changes in membership when retirements occur. (An example of the operation of the method is given in Appendix F.)

Equation (15) is somewhat unusual. In both the numerator and the denominator terms of the form \( PV_{ej} \) are summed over participants. The present values are taken at the entry date of each individual. The participants will in general have different entry dates. Thus the summation is of present values taken at different dates; apples are being added to oranges.

Equation (13) can still be used; this sharing of cost between normal costs, unfunded liability, and fund can be made to continue to function despite the unusual definition of normal cost. A result, however, is that the normal costs calculated each year are projected to be nonlevel (as either dollars or percentage of salary). The costs are nonlevel even if the assumptions are realized and despite the level dollar appearance of equation (15). This allows equation (2) or column 2 of Table 2 to be valid, but is not an acceptable practical situation.

Attempts to fit aggregate entry age normal into a consistent framework while satisfying the zero-gain criterion are explored by Tino and Sypher (1995). Their paper gives a thorough critique of the aggregate entry age normal method and finds it unacceptable.

6 Conclusions

Equation (1) indicates only that the present value of future benefits is split between the present value of future normal costs, the fund, and a balancing item, the unfunded liability. Table 2 indicates concisely the year-to-year development of the three components. All three of these components can be, and often are, varied by making changes in method and assumptions. Then the choice of cost method and asset valuation method can be made to suit the circumstances of regulation and custom.
Appendix A Benefit Allocation Methods: Accrued Liability

Under unit credit methods, it is usual to define the accrued liability \( AL \) as the present value of benefits accrued up to the valuation date, equation (A.2) below. For other cost methods the usual definition given is (A.1), \( AL^A_t = PVFB^A_t - PVFC^A_t \). It is demonstrated below that the definitions (A.1) and (A.2) are equivalent in the special case of no pre-retirement decrements. Thus, the present value of future benefits is for benefits at only one age:

\[
AL^A_t = \sum_{M_t} (PVFB^j_t - PVFC^j_t)
\]

\[
= \sum_{M_t} \frac{D_y}{D_{X_j(t)}} \bar{a}^{(12)} AB^j_y
\]

\[
- \sum_{M_t} \sum_{s=t}^{y-1} \frac{D_{X_j(s)}}{D_{X_j(t)}} \frac{D_y}{D_{X_j(s)}} \bar{a}^{(12)} (AB^j_{X_j(s)+1} - AB^j_{X_j(s)})
\]

\[
= \sum_{M_t} \frac{D_y}{D_{X_j(t)}} \bar{a}^{(12)} AB^j_y
\]

\[
- \sum_{M_t} \frac{D_y}{D_{X_j(t)}} \bar{a}^{(12)} \sum_{s=t}^{y-1} (AB^j_{X_j(s)+1} - AB^j_{X_j(s)})
\]

\[
= \sum_{M_t} \frac{D_y}{D_{X_j(t)}} \bar{a}^{(12)} AB^j_y - \sum_{M_t} \frac{D_y}{D_{X_j(t)}} \bar{a}^{(12)} (AB^j_y - AB^j_{X_j(t)})
\]
We now consider the more general case of \( n \) decrements operating in all years till the latest retirement at age \( \gamma \). We use the notation:

\[
q^{(k)}_{x_j(z)} = \text{Probability of decrement } k \text{ operating in the year of age } x_j(z) \text{ through } x_j(z+1), \text{ conditional on being a plan member at age } x_j(z).
\]

\[
L\, AB^j_{x_j(z), (x_j(s))} = \text{Portion accrued by age } x_j(s) \text{ of the lump sum equivalent of the benefit payable on decrement } k \text{ occurring in the year preceding age } x_j(z).
\]

Below it is shown that the expression of accrued liability (A.3) is equal to equation (A.4), the present value of the accrued benefit.

\[
AL_t^A = \sum_{M_t} (PVFB_t^{j,A} - PVFNC_t^{j,A}) \tag{A.3}
\]

\[
= \sum_{M_t} \sum_{z=t}^{t+y-1-x_j(t)} \sum_{k=1}^{n} \frac{D_{x_j(z)} \, q^{(k)}_{x_j(z)} \, L\, AB^j_{x_j(z+1), (x_j(z+1))}}{D_{x_j(t)}} (1+i) \\
- \sum_{M_t} \sum_{s=t}^{t+y-1-x_j(t)} \frac{D_{x_j(s)}}{D_{x_j(t)}} NC^j_{x_j(s)} \\
= \sum_{M_t} \sum_{z=t}^{t+y-1-x_j(t)} \sum_{k=1}^{n} \frac{D_{x_j(z)} \, q^{(k)}_{x_j(z)} \, L\, AB^j_{x_j(z+1), (x_j(z+1))}}{D_{x_j(t)}} (1+i) \\
- \sum_{M_t} \sum_{s=t}^{t+y-1-x_j(t)} \frac{D_{x_j(s)}}{D_{x_j(t)}} \sum_{z=s}^{t+y-1-x_j(t)} \frac{D_{x_j(z)}}{D_{x_j(s)}} \times \\
\sum_{k=1}^{n} \frac{q^{(k)}_{x_j(z)}}{1+i} \left[ L\, AB^j_{x_j(z+1), (x_j(z+1))} - L\, AB^j_{x_j(z+1), (x_j(s))} \right] \\
= \sum_{M_t} \sum_{k=1}^{n} \left\{ \sum_{z=t}^{t+y-1-x_j(t)} \frac{D_{x_j(z)} \, q^{(k)}_{x_j(z)} \, L\, AB^j_{x_j(z+1), (x_j(z+1))}}{D_{x_j(t)}} (1+i) \right\} \\
- \sum_{s=t}^{t+y-1-x_j(t)} \sum_{z=s}^{t+y-1-x_j(t)} \frac{D_{x_j(z)} \, q^{(k)}_{x_j(z)}}{D_{x_j(t)} (1+i)} \times 
\]
Appendix B Benefit Allocation Methods: Basic Funding Equations

Consider any cost method for which is valid the equation for the whole plan

$$AL_t^A = PVFB_t^A - PVFNC_t^A.$$ (B.1)

This includes the individual level cost methods because the accrued liability for those methods is defined by equation (B.1). It also includes benefit allocation methods because they too satisfy equation (B.1), as is shown in Appendix A.

Equation (B.1) is valid also for any method that satisfies

$$PVFNC_t^A + UAL_t^A + F_t^A = PVFB_t^A$$ (B.2)

and the equation

$$AL_t^A = UAL_t^A + F_t^A.$$ (B.3)

The entry age and attained age versions of the frozen initial liability method satisfy (B.2) as (B.2) is used to define their normal cost (Aitken, 1994, p. 117). Similarly the aggregate method uses (B.2) to define its
normal cost with $UAL_t$ set to zero. One could argue that (B.2) must be satisfied by any acceptable cost method. Similarly, for the frozen initial liability methods, equation (B.3) can be used to define the accrued liability (Aitken 1994, p. 117).

The aggregate method satisfies (B.3) when the definitions $UAL_t^A = 0$ and $AL_t^A = F_t^A$ are used. Thus, (B.1) is satisfied by all the usual cost methods; it is used as the usual definition and meaning of accrued (actuarial) liability.

### Appendix C Benefit Allocation Methods:
#### Reasonable Funding Method

Let us consider benefit allocation methods such as traditional unit credit and projected unit credit. Under all such methods we have, assuming that the only benefit is on normal retirement,

$$NC_t^G = \sum_{M_t} \frac{D_y^G}{D_x^G(t)} \bar{a}_y^{(12)} (AB_{X_y(t)}^{j,G} + 1 - AB_{X_y(t)}^{j,G})$$  \hfill (C.1)

where $M_t$ is the set of active members at time $t$ and $AB_{X_y(t)}^{j,G}$ is the benefit accrued up to the plan year end nearest to age $x(t)$ for member $j$. The accrued actuarial liability is consistently defined (see Appendix A) as

$$AL_t^G = \sum_{M_t} \frac{D_y^G}{D_x^G(t)} \bar{a}_y^{(12)} AB_{X_y(t)}^{j,G}$$  \hfill (C.2)

if the only benefits paid are at retirement age $y$.

Thus, expressing equation (14) as a sum over the members and noting that the basis $R(t - 1)$ used for calculating $D^R$ at time $t - 1$ is the same as $G(t)$ used for $D^G$ at time $t$, we have

$$^{tot}G_t - ^{inv}G_t$$

\[
= AL_{t-1}^R (1 + i) + NC_{t-1}^R (1 + i) - iB_{t-1} - AL_t^G \\
= \sum_{M_{t-1}} \frac{D_y^R}{D_x^R(t-1)} \bar{a}_y^{(12)} AB_{X_y(t-1)}^{j,R} (1 + i) \\
+ \sum_{M_{t-1}} \frac{D_y^R}{D_x^R(t-1)} \bar{a}_y^{(12)} (AB_{X_y(t)}^{i,R} - AB_{X_y(t-1)}^{i,R}) (1 + i) \\
- \sum_{M_{t-1}} iB_{t-1} - \sum_{M_t} \frac{D_y^G}{D_x^G(t)} \bar{a}_y^{(12)} AB_{X_y(t)}^{i,G}
\]
\[
\sum_{x(t)} \frac{D^R_y}{D^R_{x_j(t-1)}} \bar{a}^{(12)} AB_{x_j(t)}^j (1 + i) - \sum_{M_{t-1}} i B_{t-1}^j \\
- \sum_{M_t} \frac{D^G_y}{D^G_{x_j(t)}} \bar{a}^{(12)} AB_{x_j(t)}^j \\
= \sum_{M_{t-1}} \frac{D^R_y(t-1)}{D^R_{x_j(t-1)}} (1 - q^R_{x_j(t-1)}) \bar{a}^{(12)} AB_{x_j(t)}^j \\
- \sum_{M_{t-1}} \frac{D^G_{x_j(t)}}{D^G_{x_j(t)}} \bar{a}^{(12)} AB_{x_j(t)}^j \\
= 0
\]  

if the assumptions at time \( t - 1 \) are realized. The final step, equating to zero, is valid if

- \( D^R(t-1) = D^G(t) \), which is true as mentioned above.

- At all \( x(t) < y \), the set of active members \( M_t \) is \( M_{t-1} \) reduced in the proportion \( (1 - q^R_{x_j(t-1)}) \), which is true if the assumptions are realized.

- For all active members \( j \), \( AB_{x_j(t)}^j = AB_{x_j(t)}^i \), which is true if the assumptions are realized.

- For those who retire, \( i B_{t-1}^j = \bar{a}^{(12)} AB_{y}^j(t-1) \), which is true if the assumptions are realized.

Appendix D Individual Level Cost Methods:
Reasonable Funding Method

Under the individual level cost methods we have for some age \( a \) (e.g., entry if using entry age normal) for an individual \( j \):

\[
NC_t^{j,R} = \frac{PV_{a_j}FB_{x(t)}^{j,R}}{\bar{a}_{a_j;r-a_j}\mid x(t)}
\]  

where \( PV_{a_j}FB_{x(t)}^{j,R} \) is the present value at age \( a_j \) of employee \( j \)'s benefits using the revised plan. Then using the retrospective definition of
accrued liability, if no preretirement benefits are payable,

\[
AL_t^{i,R} = NC_t^{j,R} \left( \frac{\tilde{s}_{a_j; x(t)-a_j}}{\tilde{a}_{a_j;r-a_j}} \right)
\]

\[
= PV_{a_j}FB_{x(t)}^{i,R} \left( \frac{\tilde{s}_{a_j; x(t)-a_j}}{\tilde{a}_{a_j;r-a_j}} \right)
\]

\[
= PV_{x(t)}FB_{x(t)}^{i,R} \left( \frac{\tilde{a}_{a_j;x(t)-a_j}}{\tilde{a}_{a_j;r-a_j}} \right).
\]  

(D.2)  

Now equation (14) enables us to examine non-investment component of gain. We assume that the experience follows assumptions:

\[
\sum_{M_{t-1}} \left[ ^{tot} G_t - ^{inv} G_t \right]
\]

\[
= \sum_{M_{t-1}} \left( AL_t^{i,R} + NC_t^{j,R} \right) (1 + i) - \sum_{M_{t-1}} iB_{t-1}^i - \sum_{M_t} AL_t^{i,G}
\]

\[
= \sum_{M_{t-1}} \left\{ PV_{a_j}FB_{x(t-1)}^{i,R} \left[ \tilde{s}_{a_j; x_j(t)-a_j} + 1 \right] (1 + i) - iB_{t-1}^i \right\}
\]

\[
- \sum_{M_t} PV_{a_j}FB_{x(t)}^{i,G} \left( \frac{\tilde{s}_{a_j; x_j(t)-a_j}}{\tilde{a}_{a_j;r-a_j}} \right)
\]

\[
= \left[ \sum_{M_{t-1}} PV_{a_j}FB_{x(t-1)}^{i,R} (1 - q_{x_j(t-1)}) \right] - \sum_{M_t} PV_{a_j}FB_{x(t)}^{i,G} \left( \frac{\tilde{s}_{a_j; x_j(t)-a_j}}{\tilde{a}_{a_j;r-a_j}} \right)
\]

\[
- \sum_{M_{t-1}} iB_{t-1}^i
\]

\[
= 0.
\]  

(D.3)  

Again, the final step of equating to zero is valid if the assumptions are realized in the year from \( t - 1 \) to \( t \). Because the assumptions are realized, the set \( M_{t-1} \) reduced in the proportion \( (1 - q_{x_j(t-1)}) \) gives the set \( M_t \). Hence the two terms cancel in the numerator in the last stage of the above derivation. Also, assumptions \( R(t-1) \) are the same as the assumption \( G(t) \). Thus all individual level cost methods satisfy the zero-gain criterion which must be satisfied by a reasonable funding method.
Appendix E Non-Individual Methods

The frozen initial liability (entry age normal) and frozen initial liability (attained age normal) methods have, by definition, zero gain. Thus they satisfy the zero-gain criterion, though arguably through the use of a somewhat artificial procedure. These methods continue to obey equation (1) at all times because equation (1) is used to define the normal cost.

The aggregate method could be argued to give a non-zero gain by equation (10) if the actual contribution does not equal the normal cost. The subsequent forcing of the accrued liability to equal the fund is done to give the zero unfunded liability required under the aggregate method.

Appendix F An Example

Let us consider a numerical example of the operation of the aggregate entry age normal method for a two person pension plan when experience is as assumed:

<table>
<thead>
<tr>
<th>Table F.1</th>
<th>Pension Plan Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership Data</td>
<td>Member K</td>
</tr>
<tr>
<td>Date of plan inception</td>
<td>1/1/1999</td>
</tr>
<tr>
<td>Date of birth</td>
<td>1/1/1936</td>
</tr>
<tr>
<td>Date of hire</td>
<td>1/1/1981</td>
</tr>
<tr>
<td>Retirement date</td>
<td>1/1/2001</td>
</tr>
<tr>
<td>Annuity value</td>
<td>$1500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actuarial Assumptions and Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest rate:</td>
</tr>
<tr>
<td>$\hat{a}_{20}$</td>
</tr>
<tr>
<td>Pre-retirement decrements:</td>
</tr>
<tr>
<td>Method:</td>
</tr>
</tbody>
</table>

The following quantities are needed to determine the normal costs.

$$PVFB^K_{45} = \$387.628 = 1500 \times 1.07^{-20}$$
The calculation of plan normal costs from equation (15) is as follows for 1999 and 2000, where it is assumed that the actual contribution made equals the normal cost. The aggregate entry age normal method obeys the zero-gain criterion in general if equation (15) is used every year despite the resulting non-level normal cost. In other words, the experience follows the assumptions and zero gain results so the unfunded follows equation (10) with zero substituted for \( t^\text{ot} \ G \). Hence the unfunded grows only with interest: 1481.108 = 1384.213 \times 1.07 = 1293.657 \times 1.07^2.

However, the year-to-year use of equation (15) gives a non-level normal cost even if the termination and other experience is as assumed; this renders the method unacceptable. In this example the normal cost per person changes from $39.00 to $34.196.

In a practical situation it would be unacceptable also to have a negative fund after the members have both retired. The unfunded would be amortized by making amortization payments.

### Table F.2

<table>
<thead>
<tr>
<th>Jan. 1</th>
<th>Normal Cost</th>
<th>Annuity</th>
<th>Fund(^1)</th>
<th>Fund(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>78.000</td>
<td>0</td>
<td>0</td>
<td>$78.000</td>
</tr>
<tr>
<td>2000</td>
<td>34.196</td>
<td>$100</td>
<td>-$16.54</td>
<td>$17.656</td>
</tr>
<tr>
<td>2001</td>
<td>0</td>
<td>$1500</td>
<td>-$1481.108</td>
<td>-$1481.108</td>
</tr>
</tbody>
</table>

\(^1\) Fund before (2) and after (3); \(^2\) Fund after (2) and (3); Normal cost from equation (15), e.g., 78.000 = 2 \times (387.628 + 93.458)/(11.3356 + 1);

### Table F.3

Calculation of the Unfunded

<table>
<thead>
<tr>
<th>Jan. 1</th>
<th>( PVFB )</th>
<th>( PVFNC )</th>
<th>( UAL )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>$1403.616</td>
<td>$109.959</td>
<td>$1293.657</td>
</tr>
<tr>
<td>2000</td>
<td>$1401.869</td>
<td>$34.196</td>
<td>$1384.213</td>
</tr>
<tr>
<td>2001</td>
<td>$0</td>
<td>$0</td>
<td>$1481.108</td>
</tr>
</tbody>
</table>

\( UAL = \) Columns\((6) – (7) – (4)\) in Tables F.2 and F.3.
What We Say in the NAIC Annual Statement Blank Actuarial Opinion

Kenneth W. Faig, Jr.*

Abstract†

The new language adopted for the actuarial opinion in the National Association of Insurance Commissioners' model actuarial opinion and memorandum regulation has been weakened at the same time the responsibilities of the opinion actuary have been increased. The restoration of stronger language to the actuarial opinion would enhance the professional image of the actuary. If the legal environment for professional liability inhibits such a change, the opinion should be changed to describe more precisely the work performed and the conclusion reached by the actuary.

Key words and phrases: model law, valuation, professional liability, good and sufficient provision

1 A History of Life Company Actuarial Opinions

The National Association of Insurance Commissioners (NAIC) model standard valuation law as adopted in December 1990 requires that the annual statement of a life insurance company be accompanied by an

*Kenneth W. Faig, Jr., F.S.A., M.A.A.A. is the manager of actuarial services at PolySystems, Inc. His job includes providing installation, training, and client support for his firm's actuarial software. Mr. Faig has a B.A. in mathematics from Northwestern University. He was previously employed at North American Company for Life and Health Insurance and Allstate Life Insurance Co. His articles have appeared in Contingencies, Best's Reviews, the Journal of Insurance Regulation, The Actuary, and various Society of Actuaries special interest section publications. He is also proprietor of Moshassuck Press, a limited edition literary publishing house.

†The opinions expressed in this article are the author's personal opinions do not represent those of his firm or of any other person. The author thanks the anonymous reviewers whose comments helped to improve this article.
The wording of the relevant section [3(A)] of the current model law reads as follows:

Every life insurance company doing business in this state shall annually submit the opinion of a qualified actuary as to whether the reserves and related actuarial items held in support of the policies and contracts specified by the commissioner by regulation are computed appropriately, are based on assumptions which satisfy contractual provisions, are consistent with prior reported amounts and comply with applicable laws of this state. The commissioner by regulation shall define the specifics of this opinion and add any other items deemed to be necessary to its scope.

Since 1975 the annual statement instructions adopted by the NAIC have mandated the inclusion of an actuarial opinion in the annual statement filings of life insurance companies. The American Academy of Actuaries (AAA) promulgated Financial Reporting Recommendation 7 governing these statements of opinion. In June 1991 the NAIC adopted a new model actuarial opinion and memorandum regulation that provided new language for the actuarial opinion, with two different texts: one for opinions formed without asset adequacy analysis (Section 7 opinions) and one for opinions formed with asset adequacy analysis (Section 8 opinions). Asset adequacy analysis is the term adopted to indicate that the actuary has formulated his or her opinion based upon an analysis of both sides of the balance sheet, using cash flow testing or another acceptable method. The Actuarial Standards Board (ASB) followed with an Actuarial Standard of Practice (no. 22) governing Section 8 opinions and an Actuarial Compliance Guideline (no. 4) governing Section 7 opinions, in April 1993 and October 1993, respectively. Actuarial Standard of Practice no. 14, adopted by the ASB in July 1990, provides guidance to the actuary on when to perform cash flow testing.

2 The Old Actuarial Opinion Language

Amidst all the increased work that we must do to form our opinions, I wonder if we actuaries have paid enough attention to the language in

---

1See the NAIC's Model Laws, Regulations and Guidelines, (four volumes, updated to 1996). The model standard valuation law is found in volume 4, pp. 820–821. The model actuarial opinion and memorandum regulation is found at volume 4, pp. 822–824.

2The old actuarial opinion language may be found in Annual Statement Instructions: Life, Accident and Health (L/H 1994) (updated to July 28, 1994) at pp. 7–9.
which those opinions are expressed. The 1994 NAIC annual statement instructions required that the actuary opine on at least the following items:

A Aggregate reserve for life policies and contracts (Exhibit 8);
B Aggregate reserve for accident and health policies (Exhibit 9);
C Aggregate reserve for deposit funds and other liabilities without life or disability contingencies (Exhibit 10);
D Net deferred and uncollected premiums;
E Policy and contract claims—liability end of current year (Exhibit 11, part 1); and
F "Cost of collection" in excess of loading.

The model language suggested for the actuarial opinion in the 1994 NAIC annual statement instructions was as follows (emphasis added):

In my opinion the amounts carried in the balance sheet on account of the actuarial items identified above:

A Are computed in accordance with commonly accepted actuarial standards consistently applied and are fairly stated in accordance with sound actuarial principles;
B Are based on actuarial assumptions which are in accordance with or stronger than those called for in policy provisions;
C Meet the requirements of the insurance laws of (state of domicile);
D Make a good and sufficient provision for all unmatured obligations of the company guaranteed under the terms of its policies;
E Are computed on the basis of assumptions consistent with those used in computing the corresponding items in the annual statement of the preceding year-end; and
F Include provision for all actuarial reserves and related statement items which ought to be established.

In addition, the opining actuary had to indicate that the opinion was formed based on the actuarial standards of practice promulgated by the ASB.

Notice the recurrence of words and phrases with strong, positive connotations in the old opinion language: accepted actuarial standards,
consistently applied, fairly stated, sound actuarial principles, in accordance with or stronger than, meet the requirements, good and sufficient provision. The old opinion language as contained in the 1994 NAIC annual statement instructions was full of phrases with strong positive connotations. It imparted the impression that the actuary was comfortable with the company's reserve levels based upon the work he or she performed.

3 The New Actuarial Opinion Language

The purvey of the new Section 7 opinion found in the NAIC model actuarial opinion and memorandum regulation is essentially the same as that of the old opinion, except that net deferred and uncollected premiums and cost of collection in excess of loading are not explicitly mentioned (see Model Regulation 7(B)(3)). The grid accompanying the scope section for the Section 8 opinion (see Model Regulation 8(B)(2)) includes all these liabilities and, in addition, separate account liabilities, interest maintenance reserve (IMR), and asset valuation reserve (AVR).

The language of the new Section 7 opinion (see Model Regulation 7(B)(6)) is as follows (emphasis added):

In my opinion the amounts carried in the balance sheet on account of the actuarial items identified above:

A Are computed in accordance with those presently accepted actuarial standards which specifically relate to the opinion required under this section;

B Are based on actuarial assumptions which produce reserves at least as great as those called for in any contract provision as to reserve basis and method, and are in accordance with other contract provisions;

C Meet the requirements of the insurance law and regulations of the state of [state of domicile] and are at least as great as the minimum aggregate amounts required by the state in which this statement is filed;

D Are computed on the basis of assumptions consistent with those used in computing the corresponding items in the annual statement of the preceding year-end with any exceptions noted below; and

E Include provision for all actuarial reserves and related statement items which ought to be established.
The statement of conformity with ASB standards of practice is retained.

The Section 8 opinion language (see Model Regulation 8(B)(6)) follows the Section 7 language and then adds (emphasis added):

The reserves and related items, when considered in light of the assets held by the company with respect to such reserves and related actuarial items including, but not limited to, the investment earnings on such assets, and the considerations anticipated to be received and retained under such policies and contracts, make adequate provision, according to presently accepted actuarial standards of practice, for the anticipated cash flows required by the contractual obligations and related expenses of the company.

The language of the new opinion reflects some significant new obligations imposed upon the actuary. Taking assets and expenses into account is a new element of the Section 8 opinion. The requirement that the actuary specifically reference any changes in assumptions is a new element of both the Section 7 and Section 8 model opinion language. The requirement that the actuary opine regarding the aggregate compliance of the reserves with the minimum valuation standards of the state in which the statement is filed, not the domiciliary state, is also new. An American Academy of Actuaries task force chaired by Shirley Shao of the Prudential has been addressing concerns relating to state variations in valuation laws and regulations and has issued several reports to the NAIC.

4 The Language of the Old and of the New Actuarial Opinions Compared

As actuaries we should consider the impression that the new opinion language will leave with the users of life insurance company financial statements and the general public. Table 1 contrasts some of the key phrases found in the old and new forms of the opinion language:

Any practicing valuation actuary knows there are many nuances here. But practicing valuation actuaries are also readers, and virtually any reader would say that the new opinion is couched in language far weaker and far more conditioned than the corresponding language of the old opinion.

What happened? Company insolvencies happened. Lawsuits were filed against major actuarial and accounting firms that did work for
the insolvent companies. The leadership of the actuarial profession stepped up to the plate with the insurance regulatory authorities and cooperatively developed a package that included both heavier responsibilities for the valuation actuary and more protective language for the actuarial opinion. Society of Actuaries (SOA) past president Walter S. Rugland and AAA general counsel Lauren M. Bloom both worked very hard to assure that the valuation actuary was not exposed to third-party liability lawsuits as a result of the new valuation requirements. (For more on their efforts see Rugland (1992) and Bloom (1993 and 1995).) I do not take issue with the new responsibilities defined for the valuation actuary nor with the desire to protect the valuation actuary from unwarranted third-party lawsuits. I wish to address solely the question of whether the final language of the new actuarial opinion best serves these important goals.

Consider a few instances of the language changes in the actuarial opinion. We used to say we used "commonly accepted actuarial standards consistently applied." Now, we say we use "presently accepted actuarial standards which specifically relate[d] to the opinion required under this section;" I question whether this weakening of the language of the actuarial opinion is necessary to protect actuaries from lawsuits. The constitution of the United States protects us from the imposition of ex post facto law. Actuaries should be protected against the retroactive imposition of newly adopted actuarial standards of practice and actuarial compliance guidelines by a similar principle.

The old actuarial opinion said that the actuarial assumptions were "in accordance with or stronger than" those required by the policy

---

### Table 1
Comparing Opinions

<table>
<thead>
<tr>
<th>Old Opinion</th>
<th>New Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commonly accepted actuarial standards consistently applied;</td>
<td>Presently accepted actuarial standards;</td>
</tr>
<tr>
<td>Fairly stated in accordance with sound actuarial principles;</td>
<td>Specifically relate[d] to the opinion required under this section;</td>
</tr>
<tr>
<td>Assumptions which are in accordance with or stronger than;</td>
<td>Assumptions which produce reserves at least as great;</td>
</tr>
<tr>
<td>Make a good and sufficient provision for all unmatured obligations;</td>
<td>Make adequate provision for the anticipated cash flows.</td>
</tr>
</tbody>
</table>
forms. The new opinion says that the reserves are "at least as great" as those required by the policy forms. The strong, positive statement of the old opinion language has been made weak and passive. I can envision my fellow actuaries saying: "You're only talking words—the mathematics is the same!" I submit that the words leave a different flavor with readers.

Now, the real bone of contention. We used to say that the liabilities on which we were opining made "good and sufficient provision for all unmatured obligations". Now, solely in the Section 8 opinion, we say that the liabilities and the underlying assets "make adequate provision, according to presently accepted actuarial standards of practice, for the anticipated cash flows required by the contractual obligations and related expenses of the company."

It is unquestioned that the new opinion has greater breadth than the old opinion. The increased responsibility of the valuation actuary responsible for a Section 8 actuarial opinion has already been described. I question, however, whether it was necessary and prudent to go from "good and sufficient provision" to "adequate provision." Virtually every college or high school student can tell us that "good and sufficient" is higher mark than "adequate." Most students probably would tell us that "good and sufficient" represents a B grade while "adequate" represents only a C grade. If the language of the actuarial opinion needed to be weakened to this extent, perhaps we should have imitated our academic peers and converted our opinions to a pass/fail basis.

One need only consult an English language dictionary to find all the many positive qualities that the adjective "good" can denote. "Sufficient" is a more mathematical adjective with which we actuaries have greater comfort. The American Heritage Dictionary of the English Language defines "sufficient" as "as much as is needed, enough, adequate." If "sufficient" is synonymous with "adequate," why does the phrase "good and sufficient provision" leave so much more favorable an impression than the phrase "adequate provision"?

I do not believe that the colloquial usage of "good and" as an intensifier (e.g., "good and tired") is involved here. The English language recognizes the mathematical precision of adjectives such as "sufficient," "adequate," and "unique" by refusing to compare or intensify them in proper usage. I believe that "good" and "sufficient" have to be analyzed as independent and coequal modifiers of "provision." Neither is an intensifier or qualifier of the other.

Whence, then, the greater strength of "good and sufficient" as compared with "adequate"? I believe the sources are threefold:
1. The many strong positive denotations and connotations of the word "good";

2. The fact that "sufficient" (e.g., accomplishes the desired goal) has strong connotations while "adequate" (e.g., just barely accomplishes the desired goal; could have been done better) has weak connotations; and

3. The long-standing use of the phrase "good and sufficient provision" has made regulators and other users of the actuarial opinion comfortable with the language.

The third point goes further than the familiar impression left with experienced users of actuarial opinions. In the event of a dispute involving the actuarial opinion, the courts will tend to interpret established language according to established precedents. If you will, "good and sufficient provision" and the other standard opinion language become terms of art through their recognition by experienced users and interpretation by courts of law and by regulators. I believe that most potential users of the actuarial opinion would say that the use of "adequate provision" as opposed to "good and sufficient provision" has weakened the opinion.

5 Considering the Best Language for the Actuarial Opinion

I question whether the weakening of the language of the actuarial opinion is necessary to accomplish both the increased responsibility and the prudent protection of the valuation actuary. Consulting actuaries who have been involved in litigation relating to actuarial opinions which they rendered might applaud every possible dilution and conditioning of the language of the opinion. One may question whether the impact on company actuaries has been as large. Most company actuaries don't represent particularly deep pockets as far as the litigator is concerned. The officers and directors of insolvent life companies are often involved in subsequent litigation, but absent smoking gun offenses (e.g., fraudulent diversion of corporate funds), they seem to be well protected by corporate errors and omissions liability insurance coverage. If any of the chairpersons, directors, and officers of the major life companies that have become insolvent over the past two decades have become destitute as a result of the roles they played, their plight has not received coverage in the trade press.
As members of a profession, we actuaries have an obligation to our employers and to the public to perform our work in a professional manner. Public accountability is, after all, the primary element that distinguishes a profession from a trade. Every worker has an obligation to perform his or her work in a workmanlike manner although he or she does not normally issue any opinion or guarantee relating to its soundness; furthermore, his or her legal liability is lessened if he or she performs the work as a common law employee. As professionals, we have an obligation to step aside when our knowledge or our qualifications are inadequate to undertake a potential assignment. We all have to make compromises, and we are all exposed to risk. To some extent we can insure against some of the risks through professional liability insurance coverage if we or our employers can afford to do so.

I question, however, whether fear of financial liability ought to drive the form in which we express our opinions. If the legal environment relating to professional liability is sufficiently adverse, substituting a simple description of the work we perform might be a better alternative than using weak or conditioned language. Consider this proposed Section 8 opinion:

I studied tenth year surplus under the seven interest rate scenarios mandated by the NAIC model actuarial opinion and memorandum regulation. All the scenarios except the immediate 3 percent interest rate increase produced positive tenth year surplus; the immediate 3 percent interest rate increase produced $2 million negative tenth year surplus. When the company's current $10 million surplus is interjected into the study, tenth year surplus is positive under all seven interest rate scenarios. I also performed sensitivity testing as required by the model actuarial opinion and memorandum regulation. I performed my work in accordance with the actuarial standards of practice and actuarial compliance guidelines adopted by the Actuarial Standards Board.

The language is direct and tells exactly what the signatory did. I believe that such a factual description of the work performed is potentially more meaningful to users of the actuarial opinion than weak or heavily conditioned language. If the actuary had to establish additional reserves as a result of asset adequacy analysis, the opinion would state this. There are alternatives to weak or highly conditioned language. There is strong language such as "good and sufficient" that the layperson understands. Alternatively, there is an explicit summary of the
results of the actuary's work, similar to the example given in the pre­
ceding paragraph. Many actuaries would probably consider the reduc­
tion of the actuarial opinion to a mere summary of the work performed
an inadequate reflection of the opining actuary's professional respon­
sibility. In addition, a mere summary of the work performed duplicates
some of the content of the supporting actuarial memorandum.

There are undoubtedly intermediate positions between the use of
the "good and sufficient" language of the old opinion and a mere reca­
pitulation of the work performed. An early exposure draft of Actuarial
Standard of Practice no. 22 called upon the actuary to opine that the
reserves established had a "substantially better than even chance" of
providing for the company's contractual liabilities and associated ex­
penses across the range of scenarios tested. Many actuaries now believe
that reserves should allow a 20 percent to 25 percent probability of ruin
under stochastic cashflow testing, while reserves + risk-based capital
should allow for a 5 percent to 10 percent probability of ruin. Stated an­
other way, reserves should make adequate provision for the company's
contractual liabilities and expenses under moderately adverse circum­
stances, while reserves plus risk-based capital should make adequate
provision for the same liabilities and expenses under severely adverse
(but not all) circumstances.

While such results are heavily dependent upon the underlying volatil­
ity assumptions, probabilities of ruin are a concept which can be ex­
pressed meaningfully to the generally public. Perhaps another alter­
native for the actuarial opinion language is to quantify the probability
of ruin which the opining actuary believes to be inherent in the stated
reserve basis. While in the last analysis, such an opinion may be just
as subjective as opining that the reserves make "adequate" or "good
and sufficient" provision for the company's contractual liabilities and
expenses, the actuary can actually point to the calculation of the prob­
ability of ruin stated in the opinion. The problem remains that if ruin
occurs, it will in all likelihood occur under some scenario not explicitly
studied by the opining actuary.

As a profession we must decide what form of opinion best serves
the interests of our clients, our employers, ourselves, and the public.
We should not allow protection against personal financial risk to be the
predominant determinant of the language that we decide to use. We
have a duty under natural law to support ourselves and our families by
gainful work—and most of us would like to increase our wealth and to
protect it. As a profession, however, we also have a duty to our employ­
ers, to our clients, and to the public to render services in a professional
manner. These duties should come first when we consider the words that we use in our public statements of actuarial opinion.

The words we use in these statements are important. We should not change the words we use without taking into account all the many obligations that they reflect. There is an inherent danger in replacing long-established language with new language, and the danger is intensified if the new language is based on narrow professional interests. The new Section 8 opinion adds many layers of responsibility for the opining actuary. Nevertheless, the weakening and conditioning of the new opinion language gives the reader the impression that the actuary is less confident in expressing his or her opinion than before these new duties were undertaken. In fact, however, the substance and the sophistication of the work underlying the actuarial opinion are greater than ever.

I believe that the words we use in the actuarial opinion should reflect the strength of the professional work we do to form the opinion. It is unlikely that any set of future economic scenarios that we undertake to study in the process of formulating an actuarial opinion will include what actually occurs in the future. If this near-certain failure to predict the future makes it imprudent to express a professional opinion regarding reserve adequacy, I believe that reducing the opinion to a brief description of the work performed is preferable to expressing an opinion couched in weak or conditioned language.

Most good change evolves slowly, with the benefit of the wisdom garnered from experience. The potential restoration of the “good and sufficient provision” language to the actuarial opinion has been mentioned as a bargaining chip for a potential statutory reassertion of the predominant role of the domiciliary state in solvency regulation. In all regulatory processes there is inevitable give and take. With financial instruments as complex as life and health insurance and annuities, each new generation of insurance professionals must reinvent the rules in order to keep pace with change. When the pace of change is accelerating as it is today, we need to be careful when considering changes in long-established language.

6 Conclusion

I believe that any proposed revision of the NAIC model actuarial opinion and memorandum regulation ought to consider the language used to express the opinion. A thorough study of the entire issue of professional liability as it relates to the actuary would illuminate the
best direction for any future changes in the NAIC model actuarial opinion and memorandum regulation. To the greatest extent possible, the language that we use to express public statements of actuarial opinion should inspire confidence in the professional work that we have performed in forming the opinion.

The legal environment in which we live and earn our livings must remain an ever-present consideration. Any language that can be twisted to represent us as failed fortune tellers must be avoided. It would be better to describe the professional work performed than to expose ourselves to liability as failed fortune tellers.

I believe that the language we use in public statements of actuarial opinion warrants careful consideration. In the last analysis, it may be as important as the substance of the professional work that we perform.

References


---

3Readers should also note that "Actuarial Opinion" is a heading in the subject index of the book *Index to Publications of the Society of Actuaries* published from time to time by the Society of Actuaries. The *Valuation Actuary Symposium Proceedings* published annually by the Society of Actuaries, the National Association of Insurance Commissioners' *Life & Health Actuarial Subscription* (monthly), and the National Association of Insurance Commissioners' *Proceedings* (annual) are also fertile sources of information.
Constrained Forecasting of the Number of IBNR Claims

Louis G. Doray*

Abstract†

We consider the problem of forecasting the number of claims incurred. After subtracting the number of claims reported to date, the number of claims incurred but not reported (IBNR) can be forecasted. The basic model assumes that the number of claims per accident period follows an autoregressive moving average time series process. Instead of assuming the data are available in the usual claim run-off triangle format, we assume that the only data available are the number of claims reported at the valuation date for each accident interval of an observation period. Box-Jenkins methods are used to forecast the ultimate number of claims incurred and to obtain approximate confidence intervals for the number of claims incurred. The forecast of the ultimate number of claims incurred is used to derive the IBNR forecast. We show how additional information on the number of claims reported by the valuation date can be incorporated in the model when the process is autoregressive.

Key words and phrases: time series, Box-Jenkins, quadratic programming, truncated normal distribution.

*Louis G. Doray, Ph.D., A.S.A., is an assistant professor of actuarial science in the Department of Mathematics and Statistics at the University of Montreal. He obtained a B.Sc. in mathematics from the University of Montreal, an M.Math in actuarial science, and a Ph.D. in statistics from the University of Waterloo in Canada.

Dr. Doray's address is: Department of Mathematics and Statistics, University of Montreal, PO Box 6128, Station Centre-ville, Montreal PQ H3C 3J7, CANADA. Internet address: dorayl@ere.umontreal.ca

†This paper is based on part of Dr. Doray's Ph.D. dissertation (1992) at the University of Waterloo. He gratefully acknowledges the financial support received from the Natural Sciences and Engineering Research Council of Canada. Dr. Doray thanks the referees and the editor for their many suggestions that have considerably improved this paper.
1. Introduction

In all lines of insurance, there is usually a delay between the occurrence of the event giving rise to a claim and the time that the claim is first reported to the insurance company. This reporting delay, however, is more serious in certain lines of insurance than in others. For some lines, the reporting lag may be measured in days or weeks (as in the case of life insurance), while in others it may be measured in years or decades (as in the case of claims arising from environmental hazards such as asbestos).

Reinsurers experience longer delays because they have to wait for claims to exceed the retention level before being notified. Notwithstanding this reporting lag, the insurance company must estimate, at the end of each valuation year, the liability arising from the claims that have been incurred but not reported (IBNR).

Various statistical models have been proposed to estimate IBNR claims. See Van Eeghen (1981) and Taylor (1986) for a survey of some of these models. Time series models have been successfully used to model past claims amounts and forecast future claims amounts. For example, Lemaire (1982) uses an autoregressive model where the amount paid in a certain accident and development year is a linear combination of the amount in the cell above it and the one to its left. He estimates the parameters by a least-squares method. Verrall (1989) considers a similar model but uses maximum likelihood theory; his model is selected with the Akaike information criterion (AIC).

A common thread running through most research on IBNR is the assumption that the claims paid by the insurer can be grouped according to accident and development year, resulting in a triangular array of numbers. This is called the claim run-off triangle in the literature. In this paper, however, we consider a subset of the traditional structure for the data set. We assume that the only data available are the number of claims reported at the valuation date for each accident month of an observation period.

The paper is organized as follows: Section 2 reviews the basics of Box-Jenkins time series analysis. Section 3 shows how to estimate the number of IBNR claims with an ARMA\((p,q)\) model. The additional information on the number of claims reported to date is then incorporated into the model. By minimizing the sum of the squared forecast errors, subject to the ultimate number of claims incurred for a certain accident month being at least equal to the number of claims reported by the valuation date, the problem is transformed into a quadratic programming problem with linear inequality constraints. The forecasted
number of claims incurred is calculated with a modified simplex algorithm. Revised and smaller confidence intervals for the ultimate number of claims incurred with respect to each accident period can be obtained, using the truncated normal distribution. Section 4 contains an example, using actual data, of the application of the methods developed in the Section 3 to the estimation of incurred and IBNR claims and to the derivation of approximate confidence intervals for these estimates. Section 5 discusses nonstationary time series. The conclusions follow in Section 6.

2 Forecasting Using Box-Jenkins Methods

A discrete time stochastic process \( \{Z_t, t = 1, \ldots, n\} \), where \( Z_t \) takes a real valued at time \( t \), is said to be weakly stationary (see Brockwell and Davis (1991)) if:

- \( E(|Z_t^2|) < \infty \), for \( t = 1, \ldots, n \);
- \( E(Z_t) = \mu \) is constant for \( t = 1, \ldots, n \), and
- \( \text{Cov}(Z_r, Z_s) = \text{Cov}(Z_{r+t}, Z_{s+t}) \), for \( r, s, r + t, s + t = 1, \ldots, n \), i.e.,
  - the covariance structure depends only on the distance \( |r - s| \).

In time series analysis, it is usually more convenient to use the zero-mean process \( Y_t \) defined as

\[
Y_t = Z_t - \mu.
\]

In what follows, we will assume that the sequence of \( Y_t \), for \( t = 1, \ldots, n \), has a joint multivariate normal distribution. The observed time series will be represented by lower case letters \( \{z_1, \ldots, z_n\} \) and the centered observations by \( \{y_1, \ldots, y_n\} \).

2.1 Autoregressive and Moving Average Processes

There are three basic time series processes in the Box-Jenkins framework: autoregressive, moving average, and mixed autoregressive moving average processes. Given that \( \{a_t\}_{t=1}^{\infty} \) is a sequence of uncorrelated normal random variables with mean 0 and variance \( \sigma^2 \), then for \( t = 1, 2, 3, \ldots \), these processes are briefly defined below:

- The **autoregressive process** of order \( p \), \( AR(p) \), is represented as
where \( p \) is a positive integer and \( \phi_1, \ldots, \phi_p \) are constants with \( \phi_p \neq 0 \).

- The moving average process of order \( q \), \( MA(q) \), is represented by

\[
Y_t = a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}.
\]  

(2)

where \( q \) is a non-negative integer and \( \theta_1, \ldots, \theta_q \) are constants with \( \theta_q \neq 0 \). Note that

\[
\text{Var}[Y_t] = (1 + \theta_1^2 + \ldots + \theta_q^2)\sigma^2.
\]

Clearly, observations more than a distance of \( q \) steps apart are uncorrelated.

- A mixed autoregressive moving average process, \( ARMA(p, q) \), can be represented as

\[
Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}.
\]  

(3)

2.2 Model Identification

This section gives a brief overview of the process of model identification using the Box-Jenkins method. The model selected can then be used to forecast the time series. Readers unfamiliar with time series analysis using this method should consult one of the many available references on the subject (e.g., Box and Jenkins, 1976; Harvey, 1981; Abraham and Ledolter, 1983; or Pankratz, 1983).

The method consists of the following three steps:

- Identification of the process generating the data by looking at graphs of the sample autocorrelation function (a.c.f.) and partial autocorrelation function (p.a.c.f.). The sample a.c.f. is the set of autocorrelations at lag \( k \) defined by

\[
\gamma_k = \frac{\sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n} (z_t - \bar{z})^2}.
\]
The partial autocorrelation at lag $k$ is the correlation of the two residuals after regressing $y_t$ and $y_{t-k}$ on the intermediate observations $y_{t-k+1}, \ldots, y_{t-1}$.

- Estimation of the parameters of the model; and
- Verification tests to determine if the fit of the model is adequate.

For an $AR(p)$ process, the p.a.c.f. is zero after lag $p$, while the a.c.f. decays exponentially to zero. For the $MA(q)$ process, the a.c.f. cuts off after lag $q$, while the p.a.c.f. decays to zero.

After the estimation stage, the fit of the model can be checked via a test of goodness-of-fit. (See Brockwell and Davis, 1991.) The Portmanteau statistic, also called the modified Box-Pierce statistic (Box and Pierce, 1970),

\[
R = n \sum_{k=1}^{K} r_k^2(\hat{\alpha}),
\]

is calculated with $K$ usually around 20. In this formula, $r_k(\hat{\alpha})$ is the autocorrelation at lag $k$ between the residuals,

\[
\hat{\alpha}_t = y_t - \hat{y}_t
\]

\[
\hat{\alpha}_{t-k} = y_{t-k} - \hat{y}_{t-k}
\]

and $\hat{y}_t$ is the value computed with the estimated value of the parameters. When an $ARMA(p, q)$ process is appropriate, $R$ is distributed as a chi-squared random variable with $k - p - q$ degrees of freedom $(\chi^2_{k-p-q})$; large values of the test statistic $R$ indicate inadequacy of the model.

### 2.3 Forecasting

According to Anderson (1976), when the observed series of data is large, the estimation error in the parameters will not in general be serious. If we then assume that the model is known exactly for the past and that it will not change in the future, we can obtain forecasted values by minimizing the mean squared error of forecasts. Anderson (1976) shows that for the ARMA process the best $l$-step ahead forecast at time $t$, linear in $a_t$, is given by

\[
\hat{y}_t(l) = \psi_1 a_t + \psi_{l+1} a_{t-1} + \ldots
\]
where \( \psi_j \), for \( j = 1, 2, \ldots \), is the coefficient of \( x^j \) in the Taylor series expansion of

\[
\psi(x) = \frac{1 + \theta_1 x + \ldots + \theta_d x^d}{1 - \phi_1 x - \ldots - \phi_p x^p}.
\]

This forecast is unbiased and has minimum mean squared error. It therefore has minimum variance in the class of linear estimators.

Forecast errors at various leads, however, will be correlated. The \( l \)-step ahead forecast error at time \( t \) is equal to

\[
et_t(l) = Y_{t+l} - \hat{Y}_t(l) = \sum_{j=0}^{l-1} \psi_j a_{t+l-j},
\]

with variance

\[
Var[e_t(l)] = (1 + \psi_1^2 + \ldots + \psi_{l-1}^2)\sigma^2.
\]

By estimating \( \sigma^2 \) by \( \hat{\sigma}^2 \) and \( \psi_j \) by \( \hat{\psi}_j \), an approximate 100(1 - \( \alpha \))% confidence interval for the forecast is obtained:

\[
\hat{Y}_t(l) \pm [\hat{\sigma}^2 \sum_{j=0}^{l-1} \hat{\psi}_j^2]^{1/2} \xi(1 - \alpha/2),
\]  

where \( \xi(\alpha) \) satisfies

\[
\Phi(\xi(\alpha)) = \alpha \quad \text{and} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.
\]

Thus \( \xi(\alpha) \) is the 100\( \alpha \) percentile point of the standard normal distribution.

3 Estimation of IBNR

3.1 The Basic Approach

Using the theory for ARMA(\( p, q \)), we will show how it can be applied to estimate the number of IBNR claims. Let \( \{z_t, t = 1, \ldots, n\} \) be the number of claims incurred during time periods \( 1, \ldots, n \). These time intervals could be months, quarters, or years. We assume that the maximum delay between occurrence and reporting of a claim is known.

To identify the model and estimate its parameters, we will use only the data of the \( n \) periods for which the number of claims is fully known.
The process modeling the number of claims during each period can be identified by making graphs of the sample a.c.f. and p.a.c.f. The parameters of the model are then estimated by using one of the many time series software available. For an ARMA\((p, q)\) model, this would give the maximum likelihood (MLE) estimates \(\hat{\theta}_1, \ldots, \hat{\theta}_q, \hat{\phi}_1, \ldots, \hat{\phi}_p, \) their standard error, as well as the MLE for the process variance. A goodness-of-fit test to check the model adequacy is then performed.

Once the model has been validated, the forecasting of the number of claims incurred beyond time period \(n\) can be performed using equation (5). For example, the forecast at time \(n\) for periods \(n+1\) and \(n+2\) would be

\[
\hat{y}_n(1) = \hat{\psi}_1 a_n + \hat{\psi}_2 a_{n-1} + \ldots \\
\hat{y}_n(2) = \hat{\psi}_2 a_n + \hat{\psi}_3 a_{n-1} + \ldots 
\]

The standard error of the forecast is then calculated to get a confidence interval around the estimate. Let \(r_{n+1}, r_{n+2}, \ldots\) be the number of claims reported at time \(n+1, n+2, \ldots\). The number of IBNR claims predicted would then be \(\hat{y}_n(1) - r_{n+1}, \hat{y}_n(2) - r_{n+2}, \ldots\).

Note that in this subsection, we have not used the partial number of claims reported for certain periods. This is done in Section 3.2.

### 3.2 Minimum Forecast Error With Constraints

To use the additional information on the partial number of claims reported for certain periods, we will again minimize the sum of the squared forecast errors, but now subject to the constraints that each forecast should be larger than or equal to the number of claims reported by the valuation date, i.e.,

\[
\text{Minimize } \sum_{l=n+1} a_l^2, \text{ subject to } \hat{z}_l \geq r_l,
\]

where \(r_l\) is the number of claims reported for period \(l\).

This problem is a standard quadratic programming problem with linear inequality constraints. When we have an AR\((p)\) process, the problem can be rewritten in matrix form as

\[
\text{Minimize } g(\hat{y}) = \frac{1}{2} \hat{y}' Q \hat{y} \\
\text{subject to } A \hat{y} \geq b,
\]
where $Q$, $A$ are matrices and $b$, $y$ are vectors. Writing the objective function in this form will make the matrix $Q$ symmetric; it will be positive semidefinite because $g(y)$ is a sum of squares. Hillier and Lieberman (1986) or Luenberger (1984) present algorithms to solve this type of problem when the matrix $Q$ is positive semidefinite, using a modified simplex algorithm.

Note that the constraint of ultimate claim counts being no less than the number reported to date will not apply if complete salvage or subrogation recoveries are present and eliminate a previously reported claim; cumulative claim counts for a fixed accident period would then decline slightly at later evaluation dates. The method proposed here would not be applicable in this case.

### 3.3 Confidence Intervals With Constraints

Because the errors in our ARMA model are assumed to be normal, the forecasted numbers of claims for each accident period also will have a normal distribution. But that forecasted number of claims must be greater than the number reported at the end of the observation period. The forecasted number of incurred claims, therefore, will have a normal distribution truncated from below. Johnson and Kotz (1970) and Patel and Read (1982) discuss the properties of the truncated normal distribution.

A random variable $X$ has a truncated normal distribution, with lower truncation point $A$, if its pdf is given by

$$f_X(x) = \sigma^{-1} Z \left( \frac{x - \mu}{\sigma} \right) \left[ 1 - \Phi \left( \frac{A - \mu}{\sigma} \right) \right]^{-1}, \quad x \geq A$$

where

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$ 

Johnson and Kotz (1970, pp. 81-83) derive the following expressions for the expected value and the variance of $X$:

$$E(X) = \mu + \frac{\sigma Z \left( \frac{A - \mu}{\sigma} \right)}{1 - \Phi \left( \frac{A - \mu}{\sigma} \right)}$$

$$Var(X) = \sigma^2 \left[ 1 + \frac{\left( \frac{A - \mu}{\sigma} \right) Z \left( \frac{A - \mu}{\sigma} \right)}{1 - \Phi \left( \frac{A - \mu}{\sigma} \right)} - \left\{ \frac{Z \left( \frac{A - \mu}{\sigma} \right)}{1 - \Phi \left( \frac{A - \mu}{\sigma} \right)} \right\}^2 \right]$$

(7)  

(8)
The upper bound of the 95 percent confidence interval is obtained by solving for \( x \) in the equation

\[
0.95 = \frac{\Phi \left( \frac{x-\mu}{\sigma} \right) - \Phi \left( \frac{\lambda-\mu}{\sigma} \right)}{1 - \Phi \left( \frac{\lambda-\mu}{\sigma} \right)}.
\]

(9)

4 An Example

4.1 The Data

Table 1 shows the number of third party automobile liability claims reported by September 30, 1987 to a property/casualty insurance company for each accident month of the observation period January 1980 to September 1987. We assume that all the claims that occurred during accident years 1980 to 1986 have been reported by the valuation date, and because of a reporting delay, the ultimate number of claims actually incurred for each month of accident year 1987 is at least as large as the number reported (next-to-last column of Table 1). Figure 1 is a plot of the time series in Table 1.

4.2 Model Determination

We will now analyze the data of Table 1. Let \( \{z_t, t = 1, \ldots, 84\} \) represent the numbers of claims reported on September 30, 1987 for each month of the accident period 1980-1986 and let \( y_t \) be the centered observation. The graph of \( \{z_t\} \) against time (see Figure 1) shows no trend in the mean or nonconstant variance, indicating that the stationarity assumption is adequate for the data. Figure 2 contains the graphs of the sample a.c.f. and p.a.c.f.; we observe that the sample p.a.c.f. goes to zero after lag 1, suggesting the use of an AR(1) process.

Fitting that model to \( \{y_1, \ldots, y_{84}\} \) with the ITSM software,\(^1\) we obtain the MLE of \( \phi_1 \) and its estimated standard deviation (in brackets)

\[
\hat{\phi}_1 = 0.5600628 \ (0.090391).
\]

The model is therefore:

\[
y_t = 0.5600628 y_{t-1} + a_t,
\]

\(^1\)The computer program PEST, contained in the software package Interactive Time Series Modeling (ITSM) by Brockwell and Davis (1991), uses a nonlinear minimization procedure to search iteratively for the value of the parameter \( \phi_1 \) that maximizes the log-likelihood; the estimated value of \( \sigma^2 \) is then directly calculated.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JAN</td>
<td>144</td>
<td>218</td>
<td>230</td>
<td>151</td>
<td>210</td>
<td>170</td>
<td>178</td>
<td>202</td>
<td>185.86</td>
</tr>
<tr>
<td>FEB</td>
<td>149</td>
<td>243</td>
<td>179</td>
<td>135</td>
<td>142</td>
<td>177</td>
<td>130</td>
<td>156</td>
<td>165.00</td>
</tr>
<tr>
<td>MAR</td>
<td>164</td>
<td>187</td>
<td>145</td>
<td>154</td>
<td>159</td>
<td>120</td>
<td>154</td>
<td>138</td>
<td>154.71</td>
</tr>
<tr>
<td>APR</td>
<td>124</td>
<td>189</td>
<td>143</td>
<td>144</td>
<td>132</td>
<td>102</td>
<td>134</td>
<td>153</td>
<td>138.29</td>
</tr>
<tr>
<td>MAY</td>
<td>196</td>
<td>244</td>
<td>169</td>
<td>189</td>
<td>167</td>
<td>156</td>
<td>213</td>
<td>198</td>
<td>190.57</td>
</tr>
<tr>
<td>JUN</td>
<td>208</td>
<td>230</td>
<td>169</td>
<td>206</td>
<td>180</td>
<td>195</td>
<td>201</td>
<td>178</td>
<td>198.43</td>
</tr>
<tr>
<td>JUL</td>
<td>226</td>
<td>266</td>
<td>153</td>
<td>198</td>
<td>186</td>
<td>186</td>
<td>201</td>
<td>127</td>
<td>202.29</td>
</tr>
<tr>
<td>AUG</td>
<td>190</td>
<td>226</td>
<td>161</td>
<td>206</td>
<td>157</td>
<td>184</td>
<td>203</td>
<td>142</td>
<td>189.57</td>
</tr>
<tr>
<td>SEP</td>
<td>234</td>
<td>229</td>
<td>173</td>
<td>176</td>
<td>185</td>
<td>167</td>
<td>219</td>
<td>93</td>
<td>197.57</td>
</tr>
<tr>
<td>OCT</td>
<td>260</td>
<td>265</td>
<td>154</td>
<td>220</td>
<td>192</td>
<td>167</td>
<td>205</td>
<td>-</td>
<td>209.00</td>
</tr>
<tr>
<td>NOV</td>
<td>234</td>
<td>179</td>
<td>189</td>
<td>208</td>
<td>197</td>
<td>167</td>
<td>193</td>
<td>-</td>
<td>195.29</td>
</tr>
<tr>
<td>DEC</td>
<td>257</td>
<td>201</td>
<td>153</td>
<td>197</td>
<td>153</td>
<td>260</td>
<td>162</td>
<td>-</td>
<td>197.57</td>
</tr>
<tr>
<td>Mean</td>
<td>198.83</td>
<td>223.08</td>
<td>168.17</td>
<td>182</td>
<td>171.67</td>
<td>170.92</td>
<td>182.75</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
where the $a_t$s are independent $N(0, 885.562)$ random variables. An estimate of $\sigma$ is thus $\hat{\sigma} = 29.76$.

Testing the 20 residuals for randomness, we find a value of 13.6577 for the Portmanteau test statistic, which follows an asymptotic $\chi^2$ distribution with 19 degrees of freedom (the critical value at the 5 percent level is 30.1). The model therefore provides an adequate fit to the data. Figure 3 contains a residual plot; two residuals are outside the 95 percent confidence interval $[-1.96\hat{\sigma}, 1.96\hat{\sigma}]$.

4.3 Forecasting the Number of Claims Incurred

Using the above $AR(1)$ model, the next twelve monthly forecasts for 1987, as given by equation (6), appear in Table 2 along with the square root of their mean square error and 95 percent confidence interval (CI). In order to forecast, we need to assume that the average monthly number of accidents and the variance will stay the same in 1987 as in previous years.

The 95 percent confidence interval for the forecasted number of claims for an accident month in 1987 is wide and widens as the forecast...
is further in the future. It covers the actual number of claims reported at September 30, 1987 for all accident months except September (for this accident month, only the claims with a reporting lag of 0 month can be included).

In the next subsection, we see how these confidence intervals can be narrowed using the number of accidents incurred and reported in 1987 (last column of Table 2).

4.4 Minimum Forecast Error With Constraints

In Section 4.3, to forecast the number of claims for accident year 1987, we use only the number of claims that occurred during accident years 1980 to 1986. We now use the additional information on the num-
ber of claims reported for accident year 1987 to get a better forecast of the number of claims incurred in 1987.

For the AR(1) process of section 4.2, the quadratic programming problem discussed in section 3.2 becomes:

\[
\text{Minimize } \sum_{l=85}^{96} (\hat{y}_l - 0.5600628\hat{y}_{l-1})^2
\]

subject to

\[
\begin{align*}
\hat{y}_{84} &= -23.34524, & \hat{y}_{91} &\geq -58.34524, \\
\hat{y}_{85} &\geq 16.65476, & \hat{y}_{92} &\geq -43.34524, \\
\hat{y}_{86} &\geq -29.34524, & \hat{y}_{93} &\geq -92.34524, \\
\hat{y}_{87} &\geq -47.34524, & \hat{y}_{94} &\geq -185.34524, \\
\hat{y}_{88} &\geq -32.34524, & \hat{y}_{95} &\geq -185.34524, \\
\hat{y}_{89} &\geq -12.65476, & \hat{y}_{96} &\geq -185.34524, \\
\hat{y}_{90} &\geq -7.34524,
\end{align*}
\]

The figures on the right of the inequality signs represent the number of claims reported on September 30, 1987 for accident months De-
Table 2
Forecasted Numbers of Claims for 1987
Using the Box-Jenkins Method

<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast</th>
<th>(\sqrt{\text{MSE}})</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAN</td>
<td>172.27</td>
<td>29.76</td>
<td>113.94</td>
<td>230.60</td>
<td>202</td>
</tr>
<tr>
<td>FEB</td>
<td>178.02</td>
<td>34.11</td>
<td>111.16</td>
<td>244.88</td>
<td>156</td>
</tr>
<tr>
<td>MAR</td>
<td>181.24</td>
<td>35.36</td>
<td>111.93</td>
<td>250.55</td>
<td>138</td>
</tr>
<tr>
<td>APR</td>
<td>183.05</td>
<td>35.75</td>
<td>112.98</td>
<td>253.12</td>
<td>153</td>
</tr>
<tr>
<td>MAY</td>
<td>184.06</td>
<td>35.87</td>
<td>113.75</td>
<td>254.37</td>
<td>198</td>
</tr>
<tr>
<td>JUN</td>
<td>184.63</td>
<td>35.90</td>
<td>114.27</td>
<td>254.99</td>
<td>178</td>
</tr>
<tr>
<td>JUL</td>
<td>184.94</td>
<td>35.92</td>
<td>114.54</td>
<td>255.34</td>
<td>127</td>
</tr>
<tr>
<td>AUG</td>
<td>185.12</td>
<td>35.92</td>
<td>114.72</td>
<td>255.52</td>
<td>142</td>
</tr>
<tr>
<td>SEP</td>
<td>185.22</td>
<td>35.92</td>
<td>114.82</td>
<td>255.62</td>
<td>93</td>
</tr>
<tr>
<td>OCT</td>
<td>185.27</td>
<td>35.92</td>
<td>114.87</td>
<td>255.67</td>
<td>-</td>
</tr>
<tr>
<td>NOV</td>
<td>185.31</td>
<td>35.92</td>
<td>114.91</td>
<td>255.71</td>
<td>-</td>
</tr>
<tr>
<td>DEC</td>
<td>185.32</td>
<td>35.92</td>
<td>114.92</td>
<td>255.72</td>
<td>-</td>
</tr>
</tbody>
</table>

cember 1986 to December 1987 minus the average monthly number of claims for accident years 1980 to 1986 (185.34524).

In the AR(1) case, \(Q = \{Q_{ij}\}\) is a \(13 \times 13\) symmetric tridiagonal matrix where

\[
Q_{i,j} = \begin{cases} 
2\hat{\phi}_1^2 & \text{if } i = j = 1 \\
-2\hat{\phi}_1 & \text{if } i = j - 1, \text{ for } j = 2, 3, \ldots, 13 \\
-2\hat{\phi}_1 & \text{if } i = j + 1, \text{ for } j = 1, 2, \ldots, 12 \\
2(1 + \hat{\phi}_1^2) & \text{if } i = j, \text{ for } j = 2, 3, \ldots, 12 \\
2 & \text{if } i = j = 13, \text{ and} \\
0 & \text{otherwise},
\end{cases}
\]  

(10)

\(\hat{\phi}_1 = 0.560\) and \(A\) is the identity matrix.

Commercial software is available to solve quadratic programming problems of this type. Using the IMSL (1987) software,\(^2\) we obtain the forecasted values of Table 3. The forecasts for accident months Oc-

\(^2\)The IMSL subroutine QPROG uses an efficient dual algorithm in quadratic programming for a positive definite matrix \(Q\). It uses as a starting point the unconstrained minimum of the objective function and updates the solution with the Cholesky and QR factorizations.
October, November, and December 1987 are close to those produced by the Box-Jenkins method, because the constraints for those months only specify that the number of accidents should be positive.

If the only information available for accident year 1987 was that the aggregate number of claims reported on September 30, 1987 totaled 1387 (without any information on the number of claims reported for each accident month), the constraints would become

\[
\sum_{i=85}^{93} \hat{y}_i \geq -281.1716 \ (= 1387 - 9(185.34524)),
\]

\[
\hat{y}_i \geq -185.34524, \ i = 85, \ldots, 96,
\]

because all \( z_i \)'s need to be positive. The number of claims incurred for each accident month could then also be forecasted using quadratic programming.

Ordering among the number of claims to be forecasted for each accident month could also be accommodated; for example, the ordering \( \hat{y}_i \geq \hat{y}_j \geq \hat{y}_k \),

can be transformed into the two linear inequalities

\[
\hat{y}_i - \hat{y}_j \geq 0,
\]

\[
\hat{y}_j - \hat{y}_k \geq 0.
\]

In the case of an AR process of order \( p \), the matrix \( Q \) is still positive semidefinite, but becomes a band matrix.

Table 3
Forecasted Numbers of Claims Incurred For 1987 Using Quadratic Programming

<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast</th>
<th>Month</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAN</td>
<td>202.00</td>
<td>JUL</td>
<td>189.32</td>
</tr>
<tr>
<td>FEB</td>
<td>196.02</td>
<td>AUG</td>
<td>187.57</td>
</tr>
<tr>
<td>MAR</td>
<td>193.72</td>
<td>SEP</td>
<td>186.60</td>
</tr>
<tr>
<td>APR</td>
<td>194.31</td>
<td>OCT</td>
<td>186.05</td>
</tr>
<tr>
<td>MAY</td>
<td>198.00</td>
<td>NOV</td>
<td>185.74</td>
</tr>
<tr>
<td>JUN</td>
<td>192.44</td>
<td>DEC</td>
<td>185.57</td>
</tr>
<tr>
<td>Month</td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Lower 95% CI</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>JAN</td>
<td>217.63</td>
<td>13.28</td>
<td>202.00</td>
</tr>
<tr>
<td>FEB</td>
<td>192.94</td>
<td>24.75</td>
<td>156.00</td>
</tr>
<tr>
<td>MAR</td>
<td>188.75</td>
<td>29.48</td>
<td>138.00</td>
</tr>
<tr>
<td>APR</td>
<td>195.58</td>
<td>27.29</td>
<td>153.00</td>
</tr>
<tr>
<td>MAY</td>
<td>222.11</td>
<td>19.22</td>
<td>198.00</td>
</tr>
<tr>
<td>JUN</td>
<td>209.19</td>
<td>22.86</td>
<td>178.00</td>
</tr>
<tr>
<td>JUL</td>
<td>189.06</td>
<td>32.16</td>
<td>127.00</td>
</tr>
<tr>
<td>AUG</td>
<td>193.00</td>
<td>29.81</td>
<td>142.00</td>
</tr>
<tr>
<td>SEP</td>
<td>185.75</td>
<td>35.22</td>
<td>114.82</td>
</tr>
<tr>
<td>OCT</td>
<td>185.27</td>
<td>35.92</td>
<td>114.87</td>
</tr>
<tr>
<td>NOV</td>
<td>185.31</td>
<td>35.92</td>
<td>114.91</td>
</tr>
<tr>
<td>DEC</td>
<td>185.32</td>
<td>35.92</td>
<td>114.92</td>
</tr>
</tbody>
</table>

4.5 Confidence Intervals With Constraints

The 95 percent confidence intervals for the forecasted number of claims incurred for each month of accident year 1987, which appear in Table 2, are wide. Using the techniques of Section 3.3 and the number of claims reported as of September 30, 1987, they can be narrowed.

Using formulas (7) and (8), we calculate the mean and standard deviation of the forecasted number of claims for each month of 1987. The results appear in Table 4. The upper bound of the 95 percent confidence interval is obtained by solving for $x$ equation (9) and appears in the last column of Table 4.

Comparing Table 2 and Table 4, we note that for the months of October, November, and December the truncation point at 0 has no effect on the mean, the standard deviation or the confidence interval. For the other accident months, however, truncating from below reduces the standard deviation and the width of the confidence interval markedly, especially for the month of January.

Because the actuary is ultimately interested in the number of IBNR claims, the number of claims reported to date has only to be subtracted from the estimated number of claims incurred to obtain the estimated number of IBNR claims; Table 5 contains the estimated mean number of IBNR claims.
In this analysis, we consider each accident month separately. We could consider the joint multivariate distribution of the forecasted number of claims incurred for each accident month and truncate each component. This gives rise to the truncated multinormal distribution. Tallis (1961) derived the mean and variance-covariance matrix of this distribution. The calculations require the evaluation of multivariate normal integrals, not a simple task.

5 Non-Stationary Time Series

The theory presented thus far assumed that the time series was stationary, i.e. the mean of the process and the variance of the errors were constant over time. The stationarity assumption for the number of claims is not usually valid for a new line of business or a line subject to rapid growth; in such a case, it would be preferable to model claim frequency instead of counts. The assumption can be verified by plotting the observations against time. Other situations which vary from the stationary conditions can sometimes be accommodated in the Box-Jenkins method.

When the mean of the process increases linearly over time, differencing the original series will produce a new series which has a constant mean, and the theory developed previously can be applied to it. If an insurer experiences a constant growth in business, reflected in an exponentially increasing number of accidents, a logarithmic transformation of the data, followed by a differencing of the series will make it stationary. If the original data in the time series have a standard deviation which is proportional to its level, a logarithmic transformation will also make it stationary. When the inverse retransformation is used

<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast</th>
<th>Month</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAN</td>
<td>15.63</td>
<td>JUL</td>
<td>62.06</td>
</tr>
<tr>
<td>FEB</td>
<td>36.94</td>
<td>AUG</td>
<td>51.00</td>
</tr>
<tr>
<td>MAR</td>
<td>50.75</td>
<td>SEP</td>
<td>92.75</td>
</tr>
<tr>
<td>APR</td>
<td>42.58</td>
<td>OCT</td>
<td>185.27</td>
</tr>
<tr>
<td>MAY</td>
<td>24.11</td>
<td>NOV</td>
<td>185.31</td>
</tr>
<tr>
<td>JUN</td>
<td>31.19</td>
<td>DEC</td>
<td>185.32</td>
</tr>
</tbody>
</table>
to compare with the original time series, care must be taken because the forecasts will be biased. Pankratz and Dudley (1987) show how to correct for this bias.

It is conceivable that seasonality effects could affect certain lines of insurance. For example, in automobile insurance, the number of claims for damages to cars could increase during the winter months due to bad weather conditions. These seasonal models can also be incorporated in the general framework of an ARMA process by differencing corresponding monthly numbers in successive years.

To get confidence intervals for the estimates of the claims numbers, the assumption of normality of the errors was essential. Without this assumption, for a weakly stationary time series (see section 2), we can obtain least squares estimates of the parameters and best linear predictors for future values. We can not get confidence intervals based on the normal distribution or on the truncated distribution.

6 Conclusions

We have shown how to analyze the number of claims incurred in past accident periods to forecast the number for future periods. Additional information could also be taken into account to get better estimates. From these, the number of IBNR claims could be forecasted.

If the information available on the claims numbers or the claims amounts could be put into the usual claim run-off triangle format, a more traditional method of analysis such as the chain-ladder method could be employed.

References


