The Casualty Actuarial Society *Forum*

Winter 2004 Edition

Including the 2004 Ratemaking Discussion Papers and A Report of The Risk Premium Project

To CAS Members:

This is the Winter 2004 Edition of the Casualty Actuarial Society *Forum*. It contains eight Ratemaking Discussion Papers, and two additional papers. Also included is a report of the Risk Premium Project (RPP). The RPP is a team of researchers organized in response to a request for proposal distributed by the Committee on the Theory of Risk to support research on valuing property-liability risks.

The *Forum* is a nonrefereed journal printed by the Casualty Actuarial Society. The CAS is not responsible for statements or opinions expressed in the papers in this publication. See the Guide for Submission to CAS *Forum* (www.casact.org/aboutcas/forum.htm) for more information.

The CAS *Forum* is edited by the CAS Committee for the Casualty Actuarial Society *Forum*. Members of the committee invite all interested persons to submit papers on topics of interest to the actuarial community. Articles need not be written by a member of the CAS, but the paper's content must be relevant to the interests of the CAS membership. Members of the Committee for the Casualty Actuarial Society *Forum* request that the following procedures be followed when submitting an article for publication in the *Forum*:

1. Authors should submit a camera-ready original paper and two copies.
2. Authors should not number their pages.
3. All exhibits, tables, charts, and graphs should be in original format and camera-ready.
4. Authors should avoid using gray-shaded graphs, tables, or exhibits. Text and exhibits should be in solid black and white.
5. Authors should submit an electronic file of their paper using a popular word processing software (e.g., Microsoft Word and WordPerfect) for inclusion on the CAS Web Site.

The CAS *Forum* is printed periodically based on the number of call paper programs and articles submitted. The committee publishes two to four editions during each calendar year.

All comments or questions may be directed to the Committee for the Casualty Actuarial Society *Forum*.

Sincerely,

Glenn M. Walker, CAS *Forum* Chairperson

The 2004 CAS Ratemaking Discussion Papers
Presented at the
2004 CAS Ratemaking Seminar
March 11-12, 2004
Wyndham Franklin Plaza Hotel
Philadelphia, PA

The Winter 2004 Edition of the CAS *Forum* is a cooperative effort between the Committee for the CAS *Forum* and the CAS Committee on Ratemaking.

The CAS Committee on Ratemaking present for discussion eight papers prepared in response to their 2004 call for papers. This *Forum* includes papers that will be discussed by the authors at the 2004 CAS Ratemaking Seminar, March 11-12, 2004, in Philadelphia, PA.

2004 Committee on Ratemaking

John R. Pedrick, *Chairperson*

Irene K. Bass
Donald L. Closter
Beth E. Fitzgerald
Charles E. Gegax
Keith D. Holler
Eric J. Johnson
Neal Marev Leibowitz
Dennis T. McNeese
Claudine H. Modlin
Tim Query
David A. Smith
Curt A. Stewart
Jane C. Taylor
Benjamin A. Walden
Kimberley A. Ward
Jonathan White

Ratemaking Discussion Papers

Additional Papers

Report of The Risk Premium Project (In Response to the CAS Committee on Theory of Risk's Request for Research on Valuing Property-Liability Risks)

Equity Risk Premium: Expectations Great and Small

Richard A. Derrig and Elisha D. Orr

Title: Equity Risk Premium: Expectations Great and Small

Authors: Richard A. Derrig and Elisha D. Orr

Contact: Richard A. Derrig
Senior Vice President
Automobile Insurers Bureau of Massachusetts
101 Arch Street
Boston, MA 02110
617-439-4542 x718
617-439-6789 (fax)
richard@aib.org

Version 3.0
August 28, 2003

Disclaimer: The views expressed in this paper are solely the responsibility of the authors, and do not necessarily represent the views of the AIB or of any member of AIB.

Equity Risk Premium: Expectations Great and Small

What I actually think is that our prey, called the equity risk premium, is extremely elusive.

Stephen A. Ross 2001

Abstract:

The Equity Risk Premium (ERP) is an essential building block of the market value of risk. In theory, the collective action of all investors results in an equilibrium expectation for the return on the market portfolio excess of the risk-free return, the equity risk premium. The ability of the valuation actuary to choose a sensible value for the ERP, whether as a required input to CAPM valuation, or any of its descendants, is as important as choosing risk-free rates and risk relatives (betas) to the ERP for the asset at hand. The historical realized ERP for the stock market appears to be at odds with pricing theory parameters for risk aversion. Since 1985, there has been a constant stream of research, each of which reviews theories of estimating market returns, examines historical data periods, or both. Those ERP value estimates vary widely from about minus one percent to about nine percent, based on a geometric or arithmetic averaging, short or long horizons, short or long-run expectations, unconditional or conditional distributions, domestic or international data, data periods, and real or nominal returns. This paper will examine the principal strains of the recent research on the ERP and catalogue the empirical values of the ERP implied by that research. In addition, the paper will supply several time series analyses of the standard Ibbotson Associates 1926-2002 ERP data using short Treasuries for the risk-free rate. Recommendations for ERP values to use in common actuarial valuation problems will be offered.

Acknowledgments

The authors acknowledge the helpful comments from participants in the 2003 Bowles Symposium, Louise Francis, Francis Analytics & Actuarial Data Mining Inc., and four anonymous referees. The authors would also like to thank Jack Wilson for supplying his data series from Wilson & Jones (2002).

Keywords: Equity Risk Premium, Risk Premium Puzzle, Market Return Models, CAPM, Dividend Growth Models, Actuarial Valuations.

Introduction

The Equity Risk Premium (ERP) is an essential building block of the market value of risk. In theory, the collective action of all investors results in an equilibrium expectation for the return on the market portfolio excess of the risk-free return, the equity risk premium. The ability of the valuation actuary to choose a sensible value for the ERP, whether as a required input to CAPM valuation, or any of its descendants[1], is as important as choosing risk-free rates and risk relatives (betas) to the ERP for the asset at hand. Risky discount rates, asset allocation models, and project costs of capital are common actuarial uses of ERP as a benchmark rate.

The equity risk premium should be of particular interest to actuaries. For pensions and annuities backed by bonds and stocks, the actuary needs to have an understanding of the ERP and its variability compared to fixed horizon bonds. Variable products, including Guaranteed Minimum Death Benefits, require accurate projections of returns to ensure adequate future assets. With the latest research producing a relatively low equity risk premium, the rationale for including equities in insurers' asset holdings is being tested. In describing individual investment account guarantees, LaChance and Mitchell (2003) point out an underlying assumption of pension asset investing that, based only on the historical record, future equity returns will continue to outperform bonds; they clarify that those higher expected equity returns come with the additional higher risk of equity returns. Ralfe et al. (2003) support the risky equity view and discuss their pension experience with an all bond portfolio. Recent projections in some literature of a zero or negative equity risk premium challenge the assumptions underlying these views. By reviewing some of the most recent and relevant work on the issue of the equity risk premium, actuaries will have a better understanding of how these values were estimated, critical assumptions that allowed for such a low EPR, and the time period for the projection. Actuaries can then make informed decisions for expected investment results going forward.[2]

In 1985, Mehra and Prescott published their work on the so-called Equity Risk Premium Puzzle: The fact that the historical realized ERP for the stock market 1889-1978 appeared to be at odds with and, relative to Treasury bills, far in excess of asset pricing theory values based on investors with reasonable risk aversion parameters. Since then, there has been a constant stream of research, each of which reviews theories of estimating market returns, examines historical data periods, or both.[3] Those ERP value estimates vary widely from about minus one percent to about nine percent, based on geometric or arithmetic averaging, short or long horizons, short or long-run means, unconditional or conditional expectations, using domestic or international data, differing data periods, and real or nominal returns. Brealey and Myers, in the sixth edition of their standard corporate finance textbook, believe a range of 6% to 8.5% for the US ERP is reasonable for practical project valuation. Is that a fair estimate?

---

[1] The multifactor arbitrage pricing theory (APT) of Ross (1976), the three-factor model of Fama and French (1992) and the recent Mamaysky (2002) five-factor model for stocks and bonds are all examples of enhanced CAPM models.

[2] See Appendix D

[3] For example, see Cochrane (1997), Cornell (1999), or Leibowitz (2001).

Current research on the equity risk premium is plentiful (Leibowitz, 2001). This paper covers a selection of mainstream articles and books that describe different approaches to estimating the ex ante equity risk premium. We select examples of the research that cover the most important approaches to the ERP. We begin by describing the methodology of using historical returns to predict future estimates. We identify the many varieties of ERPs in order to alert the reader to the fact that numerical estimates of the ERP that appear different may instead be about the same under a common definition. We examine the well-known Ibbotson Associates 1926-2002 data series for stationarity, i.e. time invariance of the mean ERP. We show by several statistical tests that stationarity cannot be rejected and the best estimate going forward, ceteris paribus, is the realized mean. This paper will examine the principal strains of the recent research on the ERP and catalogue the empirical values of the ERP implied by that research.[4]

We first discuss how the Social Security Administration derives estimates of the equity risk premium. Then, we survey the puzzle research, that is, the literature written in response to the Equity Premium Puzzle suggested by Mehra and Prescott (1985). We cover five major approaches from the literature. Next, we report from two surveys of "experts" on the equity risk premium. Finally, after we describe the main strains of research, we explore some of the implications for practicing actuaries.

We do not discuss the important companion problem of estimating the risk relationship of an individual company, line of insurance, or project with the overall market. Within a CAPM or Fama-French framework, the problem is estimating a market beta.[5] Actuaries should be aware, however, that simple 60-month regression betas are biased low where size or non-synchronous trading is a substantial factor (Kaplan and Peterson (1998), Pratt (1998), p86). Adjustments are made to historical betas in order to remove the bias and derive more accurate estimates. Elton and Gruber (1995) explain that by testing the relationship of beta estimates over time, empirical studies have shown that an adjustment toward the mean should be made to project future betas.[6]

The Equity Risk Premium

Based on the definition in Brealey and Myers, *Principles of Corporate Finance* textbook, the equity risk premium (ERP) is the "expected additional return for making a risky investment rather than a safe one". In other words, the ERP is the difference between the market return and a risk-free return. Market returns include both dividends and capital gains. Because both the historical ERP and the prospective ERP have been referred to simply as the equity risk premium, the terms *ex post* and *ex ante* are used to differentiate between them but are often omitted. Table 1 shows the historical annual

---

[4] The research catalogued appears as Appendix B.

[5] According to CAPM, investors are compensated only for non-diversifiable, or market, risk. The market beta becomes the measurement of the extent to which returns on an individual security covary with the market. The market beta times the ERP represents the non-diversifiable expected return from an individual security.

[6] Elton and Gruber (1995), p148.

average returns from 1926-2002 for large company equities (S&P 500), Treasury Bills and Bonds, and their arithmetic differences using the Ibbotson data (Ibbotson Associates, 2003).[7]

US Equity Risk Premia 1926-2002 Annual Equity Returns and Premia versus Treasury Bills, Intermediate, and Long Term Bonds			
Horizon	**Equity Returns**	**Risk-Free Return**	**ERP**
Short	12.20%	3.83%	8.37%
Intermediate	12.20%	4.81%	7.40%
Long	12.20%	5.23%	6.97%
Source: Ibbotson Yearbook (2003)			

Table 1

In 1985, Mehra and Prescott introduced the idea of the equity risk premium puzzle. The puzzling result is that the historical realized ERP for the stock market using 1889-1978 data appeared to be at odds with and, relative to Treasury bills, far in excess of asset pricing theory values based on normal parametrizations of risk aversion. When using standard frictionless return models and historical growth rates in consumption, the real risk-free rate, and the equity risk premium, the resulting relative risk aversion parameter appears too high. By choosing a maximum relative risk aversion parameter to be 10 and using the growth in consumption, Mehra and Prescott's model produces an ERP much lower than the historical.[8] Their result inspired a stream of finance literature that attempts to solve the puzzle. Two different research threads have emerged. One thread, including behavioral finance, attempts to explain the historical returns with new models and different assumptions about investors.[9] A second thread is from a group that provides estimates of the ERP that are derived from historical data and/or standard economic models. Some in this latter group argue that historical returns may have been higher than those that should be required in the future. In a curiously asymmetric way, there are no serious studies yet concluding that the historical results are too low to serve as ex ante estimates. Although both groups have made substantial and provocative contributions, the behavioral models do not give any ex ante ERP estimates other than explaining and supporting the historical returns. We presume, until results show otherwise, the behavioralists support the historical average as the ex ante unconditional long-run expectation. Therefore, we focus on the latter to catalogue equity risk premium estimates other than the historical approach, but we will discuss both as important strains for puzzle research.

Equity Risk Premium Types

Many different types of equity risk premium estimates can be given even though they are labeled by the same general term. These estimates vary widely; currently the estimates range from about nine percent to a small negative. When ERP estimates are

---

[7] Ibbotson's 1926-2002 series from the 2003 Yearbook, Valuation Edition. The entire series is shown in Appendix A.

[8] Campbell, Lo, and MacKinlay (1997) perform a similar analysis as Mehra and Prescott and find a risk-aversion coefficient of 19, larger than the reasonable level suggested in Mehra and Prescott's paper, pp307-308.

[9] See, for example, Benartzi and Thaler (1995) and Mehra (2002).

given, one should determine the type before comparing to other estimates. We point out seven important types to look for when given an ERP estimate. They include:

- Geometric vs. arithmetic averaging
- Short vs. long investment horizon
- Short vs. long-run expectation
- Unconditional vs. conditional on some related variable
- Domestic US vs. international market data
- Data sources and periods
- Real vs. nominal returns

The average market return and ERP can be stated as a geometric or arithmetic mean return. An arithmetic mean return is a simple average of a series of returns. The geometric mean return is the compound rate of return; it is a measure of the actual average performance of a portfolio over a given time period. Arithmetic returns are the same or higher than geometric returns, so it is not appropriate to make a direct comparison between an arithmetic estimate and a geometric estimate. However, those two returns can be transformed one to the other. For example, arithmetic returns can be approximated from geometric returns by the formula.[10]

$$AR = GR + \frac{\sigma^2}{2}, \sigma^2 \text{ the variance of the (arithmetic) return process}$$

Arithmetic averages of periodic returns are to be preferred when estimating next period returns since they, not geometric averages, reproduce the proper probabilities and means of expected returns.[11] ERPs can be generated by arithmetic differences (Equity – Risk Free) or by geometric differences ([(1 + Equity)/(1 + Risk Free)]-1). Usually, the arithmetic and geometric differences produce similar estimates.[12]

A second important difference in ERP estimate types is the horizon. The horizon indicates the total investment or planning period under consideration. For estimation purposes, the horizon relates to the term or maturity of the risk-free instrument that is used to determine the ERP.[13] The Ibbotson Yearbook (2003) provides definitions for three different horizons.[14] The short-horizon expected ERP[15] is defined as "the large company stock total returns minus U.S. Treasury bill *total* returns". Note, the income return and total return are the same for U.S. Treasury bills. The intermediate-horizon expected ERP is "the large company stock total returns minus intermediate-term government bond *income* returns". Finally, the long-horizon expected ERP is "the large company stock total returns minus long-term government bond *income* returns". For the Ibbotson data, Treasury bills have a maturity of approximately one month; intermediate-term government bonds have a maturity around five years; long-term government bonds

---

[10] See Welch (2000), Dimson et al. (2002), Ibbotson and Chen (2003).

[11] For example, see Ibbotson Yearbook, Valuation Edition (2003), pp71-73 for a complete discussion of the arithmetic/geometric choice. See also Dimson et al. (2000), p35 and Brennan and Schwartz (1985).

[12] The arithmetic difference is the geometric difference multiplied by 1 + Risk Free.

[13] See Table 1.

[14] See Ibbotson 2003 Yearbook, p177.

[15] Table 1 displays the short horizon ERP calculation for the 1926-2002 Ibbotson Data.

have a maturity of about 20 years. Although the Ibbotson definitions may not apply to other research, we will classify equity risk premium estimates based on these guidelines to establish some consistency among the current research. The reader should note that Ibbotson Associates recommends the income return (or the yield) when using a bond as the risk free rate rather than the total return.[16]

A third type is the length of time of the equity risk premium forecast. We distinguish between short-run and long-run expectations. Short-run expectations refer to the current equity risk premium, or for this paper, a prediction of up to ten years. In contrast, the long-run expectation is a forecast over ten years to as much as seventy-five years for social security purposes. Ten years appears an appropriate breaking point based on the current literature surveyed.

The next difference is whether the equity risk premium estimate is unconditional or conditioned on one or more related variables. In defining this type, we refer to an admonition by Constantinides (2002, p1568) of the differences in these estimates:

> "First, I draw a sharp distinction between *conditional, short-term forecasts* of the mean equity return and premium and *estimates of the unconditional mean*. I argue that the currently low conditional short-term forecasts of the return and premium do not lessen the burden on economic theory to explain the large unconditional mean equity return and premium, as measured by their sample average over the past one hundred and thirty years."

Many of the estimates we catalogue below will be conditional ones, conditional on dividend yield, expected earnings, capital gains, or other assumptions about the future.

ERP estimates can also exhibit a US versus international market type depending upon the data used for estimation purposes and the ERP being estimated. Dimson, et al. (2002) notes that at the start of 2000, the US equity market, while dominant, was slightly less than one-half (46.1%) of the total international market for equities, capitalized at 52.7 trillion dollars. Data from the non-US equity markets are clearly different from US markets and, hence, will produce different estimates for returns and ERP.[17] Results for the entire world equity market will, of course, be a weighted average of the US and non-US estimates.

[16] The reason for this is two-fold. First, when issued, the yield is the expected market return for the entire horizon of the bond. No net capital gains are expected for the market return for the entire horizon of the bond. No capital gains are expected at the default-free maturity. Second, historical annual capital gains on long-term Government Bonds average near zero (0.4%) over the 1926-2002 period (Ibbotson Yearbook, 2003, Table 6-7).

[17] One qualitative difference can arise from the collapse of equity markets during war time.

Worldwide Equity Risk Premia, 1900-2000 Annual Equity Risk Premium Relative to Treasury Bills		
Country	**Geometric Mean**	**Arithmetic Mean**
United States	5.8%	7.7%
World	4.9%	6.2%
Source: Dimson, et al. (2002), pages 166-167		

Table 2

The next type is the data source and period used for the market and ERP estimates. Whether given an historical average of the equity risk premium or an estimate from a model using various historical data, the ERP estimate will be influenced by the length, timing, and source of the underlying data used. The time series compilations are primarily annual or monthly returns. Occasionally, daily returns are analyzed, but not for the purpose of estimating an ERP. Some researchers use as much as 200 years of history; the Ibbotson data currently uses S&P 500 returns from 1926 to the present.[18] As an example, Siegel (2002) examines a series of real US returns beginning in 1802.[19] Siegel uses three sources to obtain the data. For the first period, 1802 to 1870, characterized by stocks of financial organizations involved in banking and insurance, he cites Schwert (1990). The second period, 1871-1925, incorporates Cowles stock indexes compiled in Shiller (1989). The last period, beginning in 1926, uses CRSP data; these are the same data underlying Ibbotson Associates calculations.

Goetzmann et al. (2001) construct a NYSE data series for 1815 to 1925 to add to the 1926-1999 Ibbotson series. They conclude that the pre-1926 and post-1926 data periods show differences in both risk and reward characteristics. They highlight the fact that inclusion of pre-1926 data will generally produce lower estimates of ERPs than relying exclusively on the Ibbotson post-1926 data, similar to that shown in Appendix A. Several studies that rely on pre-1926 data, catalogued in Appendix B, show the magnitudes of these lower estimates.[20] Table 3 displays Siegel's ERPs for three subperiods. He notes that subperiod III, 1926-2001, shows a larger ERP (4.7%), or a smaller real risk-free mean (2.2%), than the prior subperiods[21].

---

[18] For the Ibbotson analysis of the small stock premium, the NYSE/AMEX/NASDAQ combined data are used with the S&P 500 data falling within deciles 1 and 3 (Ibbotson 2002 Yearbook, pp122-136.)

[19] A more recent alternative is Wilson and Jones (2002) as cited by Dimson et al. (2002), p39.

[20] Using Wilson and Jones' 1871-2002 data series, time series analyses show no significant ERP difference between the 1871-1925 period and the 1926-2002 period; one cannot distinguish the old from the new. The overall average is lower with the additional 1871-1925 data, but on a statistical basis, they are not significantly different. Assuming the equivalency of the two data series for 1871 to 1925 (series of Goetzmann et al. and Wilson & Jones), the risk difference found by Goetzmann et al. must be determined by a significantly different ERP in the pre-1871 data. The 1871-1913 return is prior to personal income tax and appears to be about 35% lower than the 1926-2002 period average of 11.8%, might reflect a zero valuation for income taxes in the pre-1914 returns. Adjusting the pre-1914 data for taxes would most likely make the ERP for the entire period (1871-2002) approximately equal to 7.5%, the 1926-2002 average.

[21] The low risk-free return is indicative of the "risk-free rate puzzle", the twin of the ERP puzzle. For details see Weil (1989).

Short-Horizon Equity Risk Premium by Subperiods			
	Subperiod I	**Subperiod II**	**Subperiod III**
	1802-1870	**1871-1925**	**1926-2001**
Real Geometric Stock Returns	7.0%	6.6%	6.9%
Real Geometric Long Term Governments	4.8%	3.7%	2.2%
Equity Risk Premium	2.2%	2.9%	4.7%
Source: Siegel (2002), pages 13 and 15.			

Table 3

Smaller subperiods will show much larger variations in equity, bill and ERP returns. Table 4 displays the Ibbotson returns and short horizon risk premia for subperiods as small as 5 years. The scatter of results is indicative of the underlying large variation (20% sd) in annual data.

Average Short-Horizon Risk Premium over Various Time Period				
		Common Stocks	U. S. Treasury Bills	Short-Horizon
Year		Total Annual Returns	Total Annual Returns	Risk Premium
All Data	1926-2002	12.20%	3.83%	8.37%
50 Year	1953-2002	12.50%	5.33%	7.17%
40 Year	1963-2002	11.80%	6.11%	5.68%
30 Year	1943-1972	14.55%	2.54%	12.02%
	1973-2002	12.21%	6.61%	5.60%
15 Year	1928-1942	5.84%	0.95%	4.89%
	1943-1957	17.14%	1.20%	15.94%
	1958-1972	11.96%	3.87%	8.09%
	1973-1987	11.42%	8.20%	3.22%
	1988-2002	13.00%	5.03%	7.97%
10 Year	1933-1942	12.88%	0.15%	12.73%
	1943-1952	17.81%	0.81%	17.00%
	1953-1962	15.29%	2.19%	13.11%
	1963-1972	10.55%	4.61%	5.94%
	1973-1982	8.67%	8.50%	0.17%
	1983-1992	16.80%	6.96%	9.84%
	1993-2002	11.17%	4.38%	6.79%
5 Year	1928-1932	- 8.25%	2.55%	-10.80%
	1933-1937	19.82%	0.22%	19.60%
	1938-1942	5.94%	0.07%	5.87%
	1943-1947	15.95%	0.37%	15.57%
	1948-1952	19.68%	1.25%	18.43%
	1953-1957	15.79%	1.97%	13.82%
	1958-1962	14.79%	2.40%	12.39%
	1963-1967	13.13%	3.91%	9.22%
	1968-1972	7.97%	5.31%	2.66%
	1973-1977	2.55%	6.19%	- 3.64%
	1978-1982	14.78%	10.81%	3.97%
	1983-1987	16.93%	7.60%	9.33%
	1988-1992	16.67%	6.33%	10.34%
	1993-1997	21.03%	4.57%	16.46%
	1998-2002	1.31%	4.18%	- 2.88%

Table 4

In calculating an expected market risk premium by averaging historical data, projecting historical data using growth models, or even conducting a survey, one must determine a proxy for the "market". Common proxies for the US market include the S&P 500, the NYSE index, and the NYSE, AMEX, and NASDAQ index.[22] For the purpose of this paper, we use the S&P 500 and its antecedents as the market. However, in the various research surveyed, many different market proxies are assumed. We have already discussed using international versus domestic data when describing different MRP types. With international data, different proxies for other country, region, or world markets are used.[23] For domestic data, different proxies have been used over time as stock market exchanges have expanded.[24] Fortunately, as shown in the Ibbotson Valuation yearbook, the issue of a US market proxy does not have a large effect on the MRP estimate because the various indices are highly correlated. For example, the S&P 500 and the NYSE have a correlation of 0.95, the S&P 500 and NYSE/AMEX/NASDAQ 0.97, and the NYSE and NYSE/AMEX/NASDAQ 0.90.[25] Therefore, the market proxy selected is one reason for slight differences in the estimates of the market risk premium.

As a final note, stock returns and risk-free rates can be stated in nominal or real terms. Nominal includes inflation; real removes inflation. The equity risk premium should not be affected by inflation because either the stock return and risk-free rate both include the effects of inflation (both stated in nominal terms) or neither have inflation (both stated in real terms). If both returns are nominal, the difference in the returns is generally assumed to remove inflation. Otherwise, both terms are real, so inflation is removed prior to finding the equity risk premium. While numerical differences in the real and nominal approaches may exist, their magnitudes are expected to be small.

Equity Risk Premia 1926-2002

As an example of the importance of knowing the types of equity risk premium estimates under consideration, Table 5 displays ERP returns that each use the same historical data, but are based on arithmetic or geometric returns and the type of horizon. The ERP estimates are quite different.[26]

[22] 2003 Ibbotson Valuation Yearbook, p92.

[23] For example, Dimson (2002) and Claus and Thomas (2001) use international market data.

[24] For a data series that is a mixture of the NYSE exchange, NYSE, AMEX, and NASDAQ stock exchange, and the Wilshire 5000, see Dimson (2002), p306.

[25] 2003 Ibbotson Valuation Yearbook, p93; using data from October 1997 to September 2002.

[26] The nominal and real ERPs are identical in Table 5 because the ERPs are calculated as arithmetic differences, and the same value of inflation will reduce the market return and the risk-free return equally. Geometric differences would produce minimally different estimates for the same types.

ERP using same historical data (1926-2002)		
RFR Description	ERP Description	ERP Historical Return
Short nominal	Arithmetic Short-horizon	8.4%
Short nominal	Geometric Short-horizon	6.4%
Short real	Arithmetic Short-horizon	8.4%
Short real	Geometric Short-horizon	6.4%
Intermediate nominal	Arithmetic Inter-horizon	7.4%
Intermediate nominal	Geometric Inter-horizon	5.4%
Intermediate real	Arithmetic Inter-horizon	7.4%
Intermediate real	Geometric Inter-horizon	5.4%
Long nominal	Arithmetic Long-horizon	7.0%
Long nominal	Geometric Long-horizon	5.0%
Long real	Arithmetic Long-horizon	7.0%
Long real	Geometric Long-horizon	5.0%

Table 5

Historical Methods

The historical methodology uses averages of past returns to forecast future returns. Different time periods may be selected, but the two most common periods arise from data provided by either Ibbotson or Siegel. The Ibbotson series begins in 1926 and is updated each year. The Siegel series begins in 1802 with the most recent compilation using returns through 2001. Appendix A provides equity risk premium estimates using Ibbotson data for the 1926-2002 period that we use in this paper for most illustrations. We begin with a look at the ERP history through a time series analysis of the Ibbotson data.

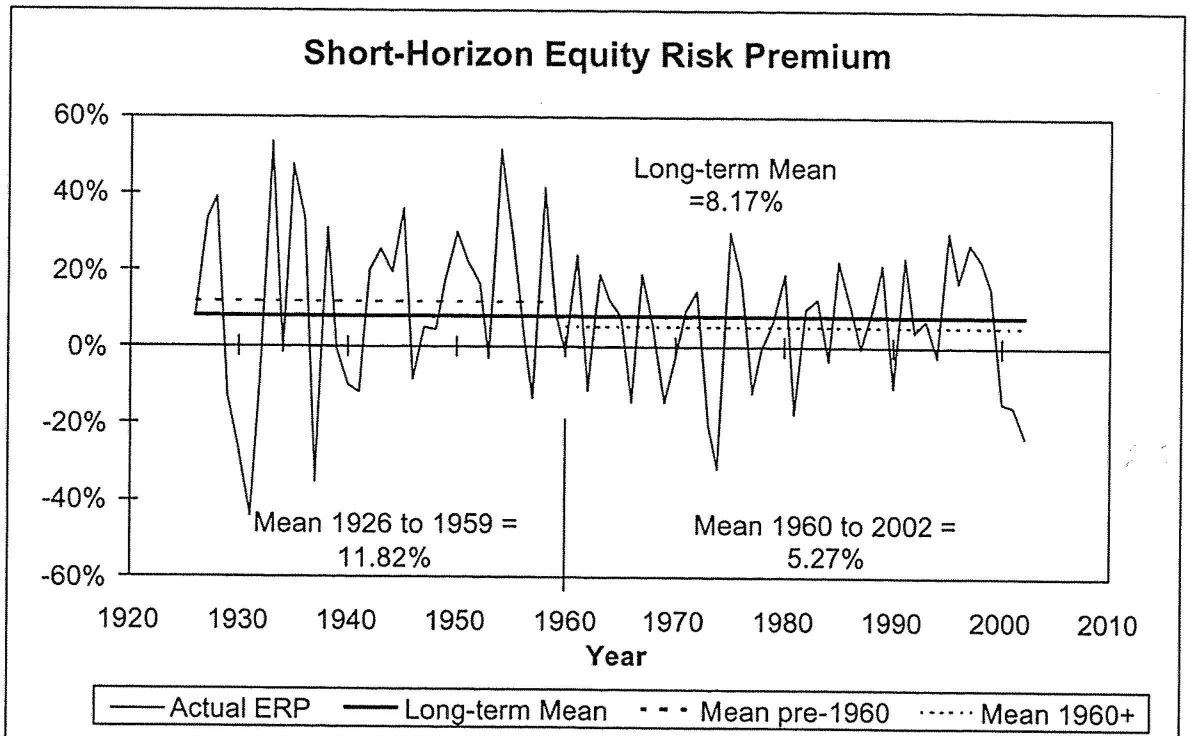
Time Series Analysis

Much of the analysis addressing the equity risk premium puzzle relies on the annual time series of market, risk-free and risk premium returns. Two opposite views can be taken of these data. One view would have the 1926-2002 Ibbotson data, or the 1802-2001 Siegel data, represent one data point; i.e., we have observed one path for the ERP through time from the many possible 77 or 200 year paths. This view rests upon the existence or assumption of a stochastic process with (possibly) inter-temporal correlations. While mathematically sophisticated, this model is particularly unhelpful without some testable hint at the details of the generating stochastic process. The practical view is that the observed returns are random samples from annual distributions that are iid, independent and identically distributed about the mean. The obvious advantage is that we have at hand 77 or 200 observations on the iid process to analyze. We adopt the latter view.

Some analyses adopt the assumption of stationarity of ERP, i.e., the true mean does not change with time. Figure 1 displays the Ibbotson ERP data and highlights two subperiods, 1926-1959 and 1960-2002.[27] While the mean ERP for the two subperiods appear quite different (11.82% vs. 5.27%), the large variance of the process (std dev 20.24%) should make them indistinguishable statistically speaking.

---

[27] The ERP shown here are the geometric differences (calculated) rather than the simple arithmetic differences in Table 1; i.e. ERP = $[(1+r_m)/(1+r_f)] - 1$. The test results are qualitatively the same for the arithmetic differences.



Source: Ibbotson Yearbook (2003), Geometric Differences

Figure 1

T-Tests

The standard T-test can be used for the null hypothesis H_0 : mean 1960-2002 = 8.17%, the 77 year mean.[28] The outcome of the test is shown in Table 6; the null hypothesis cannot be rejected.

T-Test Under the Null Hypothesis that ERP (1960-2002) = ERP (1926-2002) = 8.17%	
Sample mean 1960-2002	5.27%
Sample s.d. 1960-2002	15.83%
T value (DF=42)	-1.20
PR > |T|	0.2374
Confidence Interval 95%	(0.0040, 0.1014)
Confidence Interval 90%	(0.0121, 0.0933)

Table 6

Another T-Test can be used to test whether the subperiod means are different in the presence of unequal variances.[29] The result is similar to Table 6 and the difference of subperiod means equal to zero cannot be rejected.[30]

28 Standard statistical procedures in SAS 8.1 have been used for all tests.
29 Equality of variances is rejected at the one percent level by an F test (F=2.39, DF=33,42)
30 t-value 1.35, PR> |T| = 0.1850 with the Cochran method.

Time Trends

The supposition of stationarity of the ERP series can be supported by ANOVA regressions. The results of regressing the ERP series on time is shown in Table 7.

ERP ANOVA Regressions on Time		
Period	Time Coefficient	P-Value
1926-1959	0.004	0.355
1960-2002	0.001	0.749
1926-2002	-0.001	0.443

Table 7

There are no significant time trends in the Ibbotson ERP data.[31]

ARIMA Model

Time series analysis using the well established Box-Jenkins approach can be used to predict future series values through the lag correlation structure.[32] The SAS ARIMA procedure applied to the full 77 time series data shows:

(1) No significant autocorrelation lags.
(2) An identification of the series as white noise.
(3) ARIMA projection of year 78+ ERP is 8.17%, the 77 year average.

All of the above single time series tests point to the reasonability of the stationarity assumption for (at least) the Ibbotson ERP 77 year series.[33]

Social Security Administration

In the current debate on whether to allow private accounts that may invest in equities, the Office of the Chief Actuary of the Social Security Administration has selected certain assumptions to assess various proposals (Goss, 2001). The relevant selection is to use 7 percent as the real (geometric) annual rate of return for equities.[34] This assumption is based on the historical return of the 20th century. SSA received further support that showed the historical return for the last 200 years is consistent with this estimate, along with the Ibbotson series beginning in 1926. For SSA, the calculation of the equity risk premium uses a long-run real yield on Treasury bonds as the risk-free rate. From the assumptions in the 1995 Trustees Report, the long-run real yield on Treasury bonds that the Advisory Council proposals use is 2.3%. Using a future Treasury securities real yield of 2.3% produces a geometric equity risk premium of 4.7% over long-term Treasury securities. More recently, the Treasury securities assumption has increased to 3%[35], yielding a 4% geometric ERP over long-term Treasury securities.

---

[31] The result is confirmed by a separate Chow test on the two subperiods.
[32] See Harvey (1990), p30.
[33] The same tests applied to the Wilson and Jones 1871-2002 data series show similar results: Neither the 1871-1925 period nor the 1926-2002 period is different from the overall 1871-2002 period. The overall period and subperiods also show no trends over time.
[34] Compare Table 3, subperiod III.
[35] 1999 Social Security Trustees Report.

At the request of the Office of the Chief Actuary of the Social Security Administration (OCACT), John Campbell, Peter Diamond, and John Shoven were engaged to give their expert opinions on the assumptions Social Security mode. Each economist begins with the Social Security assumptions and then explains any difference he feels would be more appropriate.

In John Campbell's response, he considers valuation ratios as a comparison to the returns from the historical approach (Campbell 2001). The current valuation ratios are at unusual levels, with a low dividend-price ratio and high price-earnings ratio. He reasons that the prices are what have dramatically changed these ratios. Campbell presents two views as to the effect of valuation ratios in their current state. One view is that valuations will remain at the current level, suggesting much lower expected returns. The second view is a correction to the ratios, resulting in less favorable returns until the ratios readjust. He decides to give some weight to both possibilities, so he lowers the geometric equity return estimate to 5-5.5% from 7%. For the risk-free rate, he uses the yield on the long-term inflation-indexed bonds[36] of 3.5% or the OCACT assumption of 3%. Therefore, his geometric equity premium estimate is around 1.5 to 2.5%.

Peter Diamond uses the Gordon growth formula to calculate an estimate of the equity return (Diamond 2001). The classic Gordon Dividend Growth model is[37]:

$$K = (D_1 / P_0) + g$$

K = Expected Return or Discount Rate P_0 = Price this period
D_1 = Expected Dividend next period   g = Expected growth in dividends in perpetuity

Based on his analysis, he feels that the equity return assumption of 7% for the next 75 years is not consistent with a reasonable level of stock value compared to GDP. Even when increasing the GDP growth assumption, he still does not feel that the equity return is plausible. By reasoning that the next decade of returns will be lower than normal, only then is the equity return beyond that time frame consistent with the historical return. By considering the next 75 years together, he would lower the overall projected equity return to 6-6.5%. He argues that the stock market is overvalued, and a correction is required before the long-run historical return is a reasonable projection for the future. By using the OCACT assumption of 3.0% for the long-term real yield on Treasury bonds, Diamond estimates a geometric equity risk premium of about 3-3.5%.

John Shoven begins by explaining why the traditional Gordon growth model is not appropriate, and he suggests a modernized Gordon model that allows share repurchases to be included instead of only using the dividend yield and growth rate (Shoven 2001). By assuming a long-term price-earnings ratio between its current and historical value, he comes up with an estimate for the long-term real equity return of 6.125%. Using his general estimate of 6-6.5% for the equity return and the OCACT assumptions for the long-term bond yield, he projects a long-term equity risk premium of approximately 3-3.5%. All the SSA experts begin by accepting the long-run historical

36 See discussion of current yields on TIPS below.
37 Brealey and Myers (2000), p67.

ERP analyses and then modifying that by changes in the risk-free rate or by decreases in the long-term ERP based on their own personal assessments. We now turn to the major strains in ERP puzzle research.

ERP Puzzle Research

Campbell and Shiller (2001) begin with the assumption of mean reversion of dividend/price and price/earnings ratios. Next, they explain the result of prior research which finds that the dividend-price ratio predicts future prices, and historically, the price corrects the ratio when it diverts from the mean.[38] Based on this result, they then use regressions of the dividend-price ratio and the price-smoothed-earnings[39] ratio to predict future stock prices out ten years. Both regressions predict large losses in stock prices for the ten year horizon. Although Campbell and Shiller do not rerun the regression on the dividend-price ratio to incorporate share repurchases, they point out that the dividend-price ratio should be upwardly adjusted, but the adjustment only moves the ratio to the lower range of the historical fluctuations (as opposed to the mean). They conclude that the valuation ratios indicate a bear market in the near future[40]. They predict for the next ten year period negative real stock returns. They caution that because valuation ratios have changed so much from their normal level, they may not completely revert to the historical mean, but this does not change their pessimism about the next decade of stock market returns.

Arnott and Ryan (2001) take the perspective of fiduciaries, such as pension fund managers, with an investment portfolio. They begin by breaking down the historical stock returns (past 74 years since December 1925) by analyzing dividend yields and real dividend growth. They point out that the historical dividend yield is much higher than the current dividend yield of about 1.2%. They argue that the changes from stock repurchases, reinvestment, and mergers and acquisitions, which affect the lower dividend yield, can be represented by a higher dividend growth rate. However, they cap real dividend or earnings growth at the level of real economic growth. They add the dividend yield and the growth in real dividends to come up with an estimate for the future equity return; the current dividend yield of 1.2% and the economic growth rate of 2.0% add to the 3.2% estimated real stock return. This method corresponds to the dividend growth model or earnings growth model and does not take into account changing valuation levels. They cite a TIPS yield of 4.1% for the real risk-free rate return.[41] These two estimates yield a negative geometric long-horizon conditional equity risk premium.

Arnott and Bernstein (2002) begin by arguing that in 1926 investors were not expecting the realized, historical compensation that they later received from stocks. They cite bonds' reaction to inflation, increasing valuations, survivorship bias[42], and changes in

38 Campbell and Shiller (1989).
39 Earnings are "smoothed" by using ten year averages.
40 The stock market correction from year-end 1999 to year-end 2002 is a decrease of 37.6% or 14.6% per year. Presumably, the "next ten years" refers to 2000 to 2010.
41 See the current TIPS yield discussion near end of paper.
42 See Brown et al. (1992, 1995) for details on potential survivorship bias.

regulation as positive events that helped investors during this period. They only use the dividend growth model to predict a future expected return for investors. They do not agree that the earnings growth model is better than the dividend growth model both because earnings are reported using accounting methods and earnings data before 1870 are inaccurate. Even if the earnings growth model is chosen instead, they find that the earnings growth rate from 1870 only grows 0.3% faster than dividends, so their results would not change much. Because of the Modigliani-Miller theorem[43], a change in dividend policy should not change the value of the firm. They conclude that managers benefited in the "era of 'robber baron' capitalism" instead of the conclusion reached by others that the dividend growth model under-represents the value of the firm.

By holding valuations constant and using the dividend yield and real growth of dividends, Arnott and Bernstein calculate the equity return that an investor might have expected during the historical time period starting in 1802. They use an expected dividend yield of 5.0%, close to the historical average of 1810 to 2001. For the real growth of dividends, they choose the real per capita GDP growth less a reduction for entrepreneurial activity in the economy plus stock repurchases. They conclude that the net adjustment is negative, so the real GDP growth is reduced from 2.5-3% to only 1%. A fair expectation of the stock return for the historical period is close to 6.1% by adding 5.0% for the dividend yield and a net real GDP per capita growth of 1.1%. They use a TIPS yield of 3.7% for the real risk-free rate, which yields a geometric intermediate-horizon equity risk premium of 2.4% as a fair expectation for investors in the past. They consider this a "normal" equity risk premium estimate. They also opine that the current ERP is zero; i.e. they expect stocks and (risk-free) bonds to return the same amounts.

Fama and French (2002) use both the dividend growth model and the earnings growth model to investigate three periods of historical returns: 1872 to 2000, 1872 to 1950, and 1951 to 2000. Their ultimate aim is to find an unconditional equity risk premium. They cite that by assuming the dividend-price ratio and the earnings-price ratio follow a mean reversion process, the result follows that the dividend growth model or earnings growth model produce approximations of the unconditional equity return. Fama and French's analysis of the earlier period of 1872 to 1950 shows that the historical average equity return and the estimate from the dividend growth model are about the same. In contrast, they find that the 1951 to 2000 period has different estimates for returns when comparing the historical average and the growth models' estimates. The difference in the historical average and the model estimates for 1951 to 2000 is interpreted to be "unexpected capital gains" over this period. They find that the unadjusted growth model estimates of the ERP, 2.55% from the dividend model and 4.32% from the earnings model, fall short of the realized average excess return for 1951-2000. Fama and French prefer estimates from growth models instead of the historical method because of the lower standard error using the dividend growth model. Fama and French provide 3.83% as the unconditional expected equity risk premium return (referred to as the annual bias-adjusted ERP estimate) using the dividend growth model with underlying data from 1951 to 2000. They give 4.78% as the unconditional expected equity risk

[43] Brealey and Myers (2000), p447. See also discussion in Ibbotson and Chen (2003).

premium return using the earnings growth model with data from 1951 to 2000. Note that using a one-month Treasury bill instead of commercial paper for the risk-free rate would increase the ERP by about 1% to nearly 6% for the 1951-2000 period.

Ibbotson and Chen (2003) examine the historical real geometric long-run market and long risk-free returns using their "building block" methodology.[44] They use the full 1926-2000 Ibbotson Associates data and consider as building blocks all of the fundamental variables of the prior researchers. Those blocks include (not all simultaneously):

- Inflation
- Real risk-free rates (long)
- Real capital gains
- Growth of real earnings per share
- Growth of real dividends
- Growth in payout ratio (dividend/earnings)
- Growth in book value
- Growth in ROE
- Growth in price/earnings ratio
- Growth in real GDP/population
- Growth in equities excess of GDP/POP
- Reinvestment

Their calculations show that a forecast real geometric long run return of 9.4% is a reasonable extrapolation of the historical data underlying a realized 1926-2000 return of 10.7%, yielding a long horizon arithmetic ERP of 6%, or a short horizon arithmetic ERP of about 7.5%.

The authors construct six building block methods; i.e., they use combinations of historic estimates to produce an expected geometric equity return. They highlight the importance of using both dividends and capital gains by invoking the Modigliani-Miller theorem. The methods, and their component building blocks are:

- Method 1: Inflation, real risk free rate, realized ERP
- Method 2: Inflation, income, capital gains and reinvestment
- Method 3: Inflation, income, growth in price/earnings, growth in real earnings per share and reinvestment.
- Method 4: Inflation, growth rate of price/earnings, growth rate of real dividends, growth rate of payout ratio dividend yield and reinvestment
- Method 5: Inflation, income growth rate of price/earnings, growth of real book value, ROE growth and reinvestment
- Method 6: Inflation, income, growth in real GDP/POP, growth in equities excess GDP/POP and reinvestment.

[44] See Appendix D for a summary of their building block estimates. See also Pratt (1998) for a discussion of the Building Block, or Build-Up Model, cost of capital estimation method.

All six methods reproduce the historical long horizon geometric mean of 10.70% as shown in Appendix D. Since the source of most other researchers' lower ERP is the dividend yield, the authors recast the historical results in terms of ex ante forecasts for the next 75 years. Their estimate of 9.37% using supply side methods 3 and 4 is approximately 130 basis points lower than the historical result. Within their methods, they also show how the substantially lower expectation of 5.44% for the long mean geometric return is calculated by omitting one or more relevant variables. Underlying these ex ante methods are the assumptions of stationarity of the mean ERP return and market efficiency, the absence of the assumption that the market has mispriced equities. All of their methods are aimed at producing an unconditioned estimate of the ex ante ERP.

As opposed to short-run, conditional estimates from Campbell and Shiller and others, Constantinides (2002) seeks to estimate the unconditional equity risk premium, more in line with the goal of Fama and French (2002) and Ibbotson and Chen (2003). He begins with the premise that the unconditional ERP can be estimated from the historical average using the assumption that the ERP follows a stationary path. He suggests most of the other research produces conditional estimates, conditioned upon beliefs about the future paths of fundamentals such as dividend growth, price-earnings ratio and the like. While interesting in themselves, they add little to the estimation of the unconditional mean ERP.

Constantinides uses the historical return and adjusts downward by the growth in the price-earnings ratio to calculate the unconditional equity risk premium. He removes the growth in the price-earnings ratio because he is assuming no change in valuations in the unconditional state. He gives estimates using three periods. For 1872-2000, he uses the historical equity risk premium which is 6.9%, and after amortizing the growth in the price-dividend ratio or price-earnings ratio over a period as long as 129 years, the effect of the potential reduction is no change. Therefore, he finds an unconditional arithmetic, short-horizon equity risk premium of 6.9% using the 1872-2000 underlying data. For 1951-2000, he again starts with the historical equity risk premium which is 8.7% and lowers this estimate by the growth in the price-earnings ratio of 2.7% to find an unconditional arithmetic, short-horizon equity risk premium of 6.0%. For 1926-2000, he uses the historical equity risk premium which is 9.3% and reduces this estimate by the growth in the price-earnings ratio of 1.3% to find an unconditional arithmetic, short-horizon equity risk premium of 8.0%. He appeals to behavioral finance to offer explanations for such high unconditional equity risk premium estimates.

From the perspective of giving practical investor advice, Malkiel (1999) discusses "the age of the millennium" to give some indication of what investors might expect for the future. He specifically estimates a reasonable expectation for the first few decades of the twenty-first century. He estimates the future bond returns by giving estimates if bonds are held to maturity with corporate bonds of 6.5-7%, long-term zero-coupon Treasury bonds of about 5.25%, and TIPS with a 3.75% return. Depending on the desired level of risk, Malkiel indicates bondholders should be more favorably compensated in the future compared to the historical returns from 1926 to 1998. Malkiel uses the earnings growth model to predict future equity returns. He uses the current dividend yield of 1.5% and an earnings growth estimate of 6.5%, yielding an 8% equity return estimate compared with an 11% historical return. Malkiel's estimated range of the equity risk premium is from 1% to 4.25%, depending on the risk-free instrument selected. Although his equity risk premium is lower than the historical return, his selection of a relatively high earnings growth rate is similar to Ibbotson and Chen's forecasted models. In contrast with Ibbotson and Chen, Malkiel allows for a changing equity risk premium and advises investors to not rely solely on the past "age of exuberance" as a guide for the future. Malkiel points out the impact of changes in valuation ratios, but he does not attempt to predict future valuation levels.

Finally, Mehra (2002) summarizes the results of the research since the ERP puzzle was posed. The essence of the puzzle is the inconsistency of the ERPs produced by descriptive and prescriptive economic models of asset pricing on the one hand and the historical ERPs realized in the US market on the other. Mehra and Prescott (1985) speculated that the inconsistency could arise from the inadequacy of standard models to incorporate market imperfections and transaction costs. Failure of the models to reflect reality rather than failure of the market to follow the theory seems to be Mehra's conclusion as of 2002. Mehra points to two promising threads of model-modifying research. Campbell and Cochrane (1999) incorporate economic cycles and changing risk aversion while Constantinides et al. (2002) propose a life cycle investing modification, replacing the representative agent by segmenting investors into young, middle aged, and older cohorts. Mehra sums up by offering:

> "Before we dismiss the premium, we not only need to have an understanding of the observed phenomena but also why the future is likely to be different. In the absence of this, we can make the following claim based on what we know. Over the long horizon the equity premium is likely to be similar to what it has been in the past and the returns to investment in equity will continue to substantially dominate those in bonds for investors with a long planning horizon."

Financial Analyst Estimates

Claus and Thomas (2001) and Harris and Marston (2001) both provide equity premium estimates using financial analysts' forecasts. However, their results are rather different. Claus and Thomas use an abnormal earnings model with data from 1985 to 1998 to calculate an equity risk premium as opposed to using the more common dividend growth model. Financial analysts project five year estimates of future earnings growth rates. When using this five year growth rate for the dividend growth rate in perpetuity in the Gordon growth model, Claus and Thomas explain that there is a potential upward bias in estimates for the equity risk premium. Therefore, they choose to use the abnormal earnings model instead and only let earnings grow at the level of inflation after five years. The abnormal earnings model replaces dividends with "abnormal earnings" and discounts each flow separately instead of using a perpetuity. The average estimate that they find is 3.39% for the equity risk premium. Although it is generally recognized that financial analysts' estimates have an upward bias, Claus and Thomas propose that in the current literature, financial analysts' forecasts have underestimated short-term earnings in order for management to achieve earnings estimates in the slower economy. Claus and Thomas conclude that their findings of the ERP using data from the past fifteen years are not in line with historical values.

Harris and Marston use the dividend growth model with data from 1982 to 1998. They assume that the dividend growth rate should correspond to investor expectations. By using financial analysts' longest estimates (five years) of earnings growth in the model, they attempt to estimate these expectations. They argue that if investors are in accord with the optimism shown in analysts' estimates, even biased estimates do not pose a drawback because these market sentiments will be reflected in actual returns. Harris and Marston find an equity risk premium estimate of 7.14%. They find fluctuations in the equity risk premium over time. Because their estimates are close to historical returns, they contend that investors continue to require a high equity risk premium.

Survey Methods

One method to estimate the ex ante equity risk premium is to find the consensus view of experts. John Graham and Campbell Harvey perform a survey of Chief Financial Officers to determine the average cost of capital used by firms. Ivo Welch surveys financial economists to determine the equity risk premium that academic experts in this area would estimate.

Graham and Harvey administer surveys from the second quarter of 2000 to the third quarter of 2002 (Graham and Harvey, 2002). For their survey format, they show the current ten year bond yield and then ask CFOs to provide their estimate of the S&P 500 return for the next year and over the next ten years. CFOs are actively involved in setting a company's individual hurdle[45] rate and are therefore considered knowledgeable about investors' expectations.[46] When comparing the survey responses of the one and ten year returns, the one year returns have so much volatility that they conclude that the ten-year equity risk premium is the more important and appropriate return of the two when making financial decisions such as hurdle rates and estimating cost of capital. The average ten-year equity risk premium estimate varies from 3% to 4.7%.

The most current Welch survey compiles the consensus view of about five hundred financial economists (Welch 2001). The average arithmetic estimate for the 30-year equity risk premium relative to Treasury bills is 5.5%; the one-year arithmetic equity risk premium consensus is 3.4%. Welch deduces from the average 30-year geometric

[45] A "hurdle" rate is a benchmark cost of capital used to evaluate projects to accept (expected returns greater than hurdle rate) or reject (expected returns less than hurdle rate).
[46] Graham and Harvey claim three-fourths of the CFOs use CAPM to estimate hurdle rates.

equity return estimate of 9.1% that the arithmetic equity return forecast is approximately 10%.[47]

Welch's survey question allows the participants to self select into different categories based upon their knowledge of ERP. The results indicate that the responses of the less ERP knowledgeable participants showed more pessimism than those of the self reported experts. The experts gave 30-year estimates that are 30 to 150 basis points above the estimates of the non-expert group.

Differences in Forecasts across Expertise Level				
Relative Expertise	**Statistic**	**Stock Market**	**Equity Premium**	
		30-Year Geometric	30-Year Arithmetic	30-Year Geometric
188 Less Involved	Mean	8.5%	4.9%	4.4%
	Median	8%	5%	4%
	IQ Range	6%-10%	3%-6%	2%-5.5%
235 Average	Mean	9.2%	5.8%	4.8%
	Median	9%	5%	4%
	IQ Range	7.5%-10%	3.5%-7%	3%-6%
72 Experts	Mean	10.1%	6.2%	5.4%
	Median	9%	5.4%	5%
	IQ Range	8%-11%	4%-7.5%	3.4%-6%
Data Source: Welch (2001), Table 5				

Table 8

Table 8 shows that there may be a "lemming" effect, especially among economists who are not directly involved in the ERP question. Stated differently, all the academic and popular press, together with the prior Welch survey[48] could condition the non-expert, the "less involved", that the expected ERP was lower than historic levels.

The Behavioral Approach

Benartzi and Thaler (1995) analyze the equity risk premium puzzle from the point of view of prospect theory (Kahneman and Tversky; 1979). Prospect theory[49] has "loss aversion", the fact that individuals are more sensitive to potential loss than gain, as one of its central tenets. Once an asymmetry in risk aversion is introduced into the model of the rational representative investor or agent, the unusual risk aversion problem raised initially by Mehra and Prescott (1985) can be "explained" within this behavioral model of decision-making under uncertainty. Stated differently, given the historical ERP series, there exists a model of investor behavior that can produce those or similar results. Benartzi and Thaler combine loss aversion with "mental accounting", the behavioral process people use to evaluate their status relative to gains and losses compared to expectations, utility and wealth, to get "myopic loss aversion". In particular, mental

[47] For the Ibbotson 1926-2002 data, the arithmetic return is about 190 basis points higher than the geometric return rather than the inferred 90 basis points. This suggests the participant's beliefs may not be internally consistent.

[48] The prior Welch survey in 1998 had a consensus ERP of about 7%.

[49] A current survey of the applications of prospect theory to finance can be found in Benartzi et al. (2001).

accounting for a portfolio needs to take place infrequently because of loss aversion, in order to reduce the chances of observing loss versus gain. The authors concede that there is a puzzle with the standard expected utility-maximizing paradigm but that the myopic loss aversion view may resolve the puzzle. The authors' views are not free of controversy; any progress along those lines is sure to match the advance of behavioral economics in the large.

The adoption of other behavioral aspects of investing may also provide support for the historical patterns of ERPs we see from 1802-2002. For example, as the true nature of risk and rewards has been uncovered by the virtual army of 20th century researchers, and as institutional investors held sway in the latter fifty years of the century, the demand for higher rewards seen in the later historical data may be a natural and rational response to the new and expanded information set. Dimson et al. (2002, Figure 4-6) displays increasing real US equity returns of 6.7, 7.4, 8.2 and 10.2 for periods of 101, 75, 50 and 25 years ending in 2001 consistent with this "risk-learning" view.

Next Ten Years

The "next ten years" is an issue that experts reviewing Social Security assumptions and Campbell and Shiller address either explicitly or implicitly. Experts evaluating Social Security's proposals predicted that the "next ten years", indicating a period beginning around 2000, of returns were likely to be below the historical return. However, a historical return was recommended as appropriate for the remaining 65 of the 75 years to be projected. For Campbell and Shiller (2001), the period they discuss is approximately 2000-2010. Based on the current state of valuation ratios, they predict lower stock market returns over "the next ten years". These expert predictions, and other pessimistic low estimates, have already come to fruition as market results 2000 through 2002.[50] The US equities market has decreased 37.6% since 1999, or an annual decrease of 14.6%. Although these forecasts have proved to be accurate in the short term, for future long-run projections, the market is not at the same valuation today as it was when these conditional estimates were originally given. Therefore, actuaries should be wary of using the low long-run estimates made prior to the large market correction of 2000-2002.

Treasury Inflation Protection Securities (TIPS)

Several of the ERP researchers refer to TIPS when considering the real risk free rates. Historically, they adjust Treasury yields downward to a real rate by an estimate of inflation, presumably for the term of the Treasury security. As Table 3 shows using the Siegel data, the modern era data show a low real long-term risk-free rate of return (2.2%). This contrasts with the initial[51] TIPS issue yields of 3.375%. Some researchers use those TIPS yields as (market) forecasts of real risk-free returns for intermediate and long-horizon, together with reduced (real) equity returns to produce low estimates of ex ante ERPs. None consider the volatility of TIPS as indicative of the accuracy of their ERP estimate.

[50] The Social Security Advisory Board will revisit the seventy five year rate of return assumption during 2003, Social Security Advisory Board (2002).
[51] TIPS were introduced by the Treasury in 1996 with the first issue in January, 1997.

Table 9 shows a recent market valuation of ten and thirty year TIPS issued in 1998-2002.

Inflation-Indexed Treasury Securities		
Maturity	Coupon Issue Rate	Yield to Maturity
1/11	3.500	1.763
1/12	3.375	1.831
7/12	3.000	1.878
4/28	3.625	2.498
4/29	3.875	2.490
4/32	3.375	2.408
Source: WSJ 1 2/24/2003		

Table 9

Note the large 90-180 basis point decrease in the current "real" yields from the issue yields as recent as ten months ago. While there can be several explanations for the change (revaluation of the inflation option, flight to Treasury quality, paucity of 30 year Treasuries), the use of these current "real" risk free yields, with fixed expected returns, would raise ERPs by at least one percent.

Conclusion

This paper has sought to bring the essence of recent research on the equity risk premium to practicing actuaries. The researchers covered here face the same ubiquitous problems that actuaries face daily: Do I rely on past data to forecast the future (costs, premiums, investments) or do I analyze the past and apply informed judgment as to future differences, if any, to arrive at actuarially fair forecasts? Most of the ERP estimates lower than the unconditional historical estimate have an undue reliance on recent lower dividend yields (without a recognition of capital gains[52]) and/or on data prior to 1926.

Despite a spate of research suggesting ex ante ERPs lower than recent realized ERPs, actuaries should be aware of the range of estimates covered here (Appendix B); be aware of the underlying assumptions, data and terminology; and be aware that their independent analysis is required before adopting an estimate other than the historical average. We believe that the Ibbotson-Chen (2003) layout, reproduced here as Appendix D, offers the actuary both an understanding of the fundamental components of the historical ERP and the opportunity to change the estimates based upon good judgment and supportable beliefs. We believe that reliance solely on "expert" survey averages, whether of financial analysts, academic economists, or CFOs, is fraught with risks of statistical bias to fair estimates of the forward ERP.

---

[52] Under the current US tax code, capital gains are tax-advantaged relative to dividend income for the vast majority of equity holders (households and mutual funds are 55% of the total equity holders, Federal Flow of Funds, 2002 Q3, Table L-213). Curiously, the reverse is true for property-liability insurers because of the 70% stock dividend exclusion afforded insurers.

It is dangerous for actuaries to engage in simplistic analyses of historical ERPs to generate ex ante forecasts that differ from the realized mean.[53] The research we have catalogued in Appendix B, the common level ERPs estimated in Appendix C, and the building block (historical) approach of Ibboston and Chen in Appendix D all discuss important concepts related to both ex post and ex ante ERPs and cannot be ignored in reaching an informed estimate. For example, Richard Wendt, writing in a 2002 issue of Risks and Rewards, a newsletter of the Society of Actuaries, concludes that a linear relationship is a better predictor of future returns than a "constant" ERP based on the average historical return. He arrives at this conclusion by estimating a regression equation[54] relating long bond yields with 15-year geometric mean market returns starting monthly in 1960. First, there is no significant relationship between short, intermediate or long-term income returns over 1926-2002 (or 1960-2002) and ERPs, as evidenced by simple regressions using Ibbotson data.[55] Second, if the linear structural equation indeed held, there would be no need for an ERP since the (15-year) return could be predicted within small error bars. Third, there is always a negative bias introduced when geometric averages are used as dependent variables (Brennan and Schwartz, 1985). Finally, the results are likely to be spurious due to the high autocorrelations of the target and independent variables; an autocorrelation correction would eliminate any significant relationship of long-yields to the ERP.

Actuaries should also be aware of the variability of both the ERP and risk-free rate estimates discussed in this paper (see Tables 4 and 9). All too often, return estimates are made without noting the error bars and that can lead to unexpected "surprises". As one example, recent research by Francis Longstaff (2002), proposes that a 1991-2001 "flight to quality" has created a valuation premium (and lowered yields) in the entire yield curve of Treasuries. He finds a 10 to 16 basis point liquidity premium throughout the zero coupon Treasury yield curve. He translates that into a 10% to 15% pricing difference at the long end. This would imply a simple CAPM market estimate for the long horizon might be biased low.

Finally, actuaries should know that the research catalogued in Appendix B is not definitive. No simple model of ERP estimation has been universally accepted. Undoubtedly, there will be still more empirical and theoretical research into this data rich financial topic. We await the potential advances in understanding the return process that the behavioral view may uncover.

Post Script: Appendices A-D

We provide four appendices that catalogue the ERP approaches and estimates discussed in the paper. Actuaries, in particular, should find the numerical values, and descriptions of assumptions underlying those values, helpful for valuation work that

[53] ERPs are derived from historical or expected after corporate tax returns. Pre-tax returns depend uniquely on the tax schedule for the differing sources of income.

[54] 15-year mean returns = 2.032 (Long Government Bond Yield) – 0.0242, $R^2 = 0.882$.

[55] The p-values on the yield-variables in an ERP/Yield regression using 1926-2002 annual data are 0.1324, 0.2246, and 0.3604 for short, intermediate and long term yields respectively with adjusted r square virtually zero.

adjusts for risk. Appendix A provides the annual Ibbotson data from 1926 through 2002 from Ibbotson Associates referred to throughout this paper. The equity risk-premium shown is a simple difference of the arithmetic stock returns and the arithmetic U.S. Treasury Bills total returns. Appendix B is a compilation of articles and books related to the equity risk premium. The puzzle research section contains the articles and books that were most related to addressing the equity risk premium puzzle. Page 1 of Appendix B gives each source, along with risk-free rate and equity risk premium estimates. Then, each source's estimate is classified by type (indicated with an X for the appropriate type). Page 2 of Appendix B shows further details collected from each source. This page adds the data period used, if applicable, and the projection period. We also list the general methodology used in the reference. The final three pages of Appendix B provide the footnotes which give additional details on the sources' intent.

Appendix C adjusts all the equity risk premium estimates to a short-horizon, arithmetic, unconditional ERP estimate. We begin with the authors' estimates for a stock return (the risk-free rate plus the ERP estimate). Next, we make adjustments if the ERP "type" given by the author(s) is not given in this format. For example, to adjust from a geometric to an arithmetic ERP estimate, we adjust upwards by the 1926-2002 historical difference in the arithmetic large company stocks' total return and the geometric large company stocks' total return of 2%. Next, if the estimate is given in real instead of nominal terms, we adjust the stock return estimate upwards by 3.1%, the 1926-2002 historical return for inflation.

We make an approximate adjustment to move the estimate from a conditional to unconditional estimate based on Fama and French (2002). Using the results for the 1951-2000 period shown in Table 4 of their paper and the standard deviations provided in Table 1, we have four adjustments based on their data. For the 1951-2000 period, Fama and French use an adjustment of 1.28% for the dividend growth model and 0.46% for the earnings growth model. Following a similar calculation, the 1872-2000 period would require a 0.82% adjustment using a dividend growth model; the 1872-1950 period would require a 0.54% adjustment using a dividend growth model. Earnings growth models were used by Fama and French only for the 1951-2000 data period. Therefore, we selected the lowest adjustment (0.46%) as a minimum adjustment from a conditional estimate to an unconditional estimate. Finally, we subtract the 1926-2002 historical U.S. Treasury Bills' total return to arrive at an adjusted equity risk premium.

These adjustments are only approximations because the various sources rely on different underlying data, but the changes in the ERP estimate should reflect the underlying concept that different "types" of ERPs cannot be directly compared and require some attempt to normalize the various estimates.

Page 1 of Appendix D is a table from Ibbotson and Chen which breaks down historical returns using various methods that correspond to their 2003 paper (reprinted with permission of Ibbotson Associates). The bottom portion provides forward-looking estimates. Page 2 of Appendix D is provided to show the formulas that Ibbotson and Chen develop within their paper.

References

Arnott, Robert D., and Peter L. Bernstein, 2002, What Risk Premium Is "Normal"?. *Financial Analysts Journal,* 58, 64-85.

Arnott, Robert D., and Ronald J. Ryan, 2001, The Death of the Risk Premium: Consequences of the 1990s, *Journal of Portfolio Management,* 27, 61-74.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect Theory and Asset Prices, *Quarterly Journal of Economics,* 116, 1-53.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic Loss Aversion and the Equity Premium Puzzle, *Quarterly Journal of Economics,* 110, 73-92.

Brealey, Richard A., and Stewart C. Myers, 2000, *Principles of Corporate Finance,* 6th ed. (Boston: Irwin McGraw-Hill).

Brennan, Michael J., and Eduardo S. Schwartz, 1985, On the Geometric Mean Index: A note, *Journal of Financial and Quantitative Analysis*, 20,119-122.

Brown, Stephen J., William Goetzmann, Roger G. Ibboston and Stephen A. Ross 1992, Survivorship Bias in Performance Studies, *Review of Financial Studies,* 5, 553-580.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance,* 50, 853-873.

Campbell, John Y., 2001, Forecasting U.S. Equity Returns in the 21st Century, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Campbell, John Y., and John H. Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economics* 107: 205-251.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton: Princeton University Press).

Campbell, John Y., and Robert J. Shiller, 1989, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-228.

Campbell, John Y., and Robert J. Shiller, 2001, Valuation Ratios and the Long-Run Stock Market Outlook: An Update, Working paper #8221, NBER, forthcoming in

Advances in Behavioral Finance, Volume II, Nicholas Barberis and Richard Thaler eds., Russell Sage Foundation, 2003.

Claus, James, and Jacob Thomas, 2001, Equity Premia as Low as Three Percent?, *Journal of Finance,* 56, 1629-1666.

Cochrane, John H., 1997, Where is the Market Going? Uncertain Facts and Novel Theories, *Economic Perspectives,* 21, 3-37.

Constantinides, George M., John B. Donaldson, and Rajnish Mehra, 2002, Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle, *Quarterly Journal of Economics* 117: 269-296.

Constantinides, George M., 2002, Rational Asset Prices, *Journal of Finance,* 57, 1567-1591.

Cornell, Bradford, 1999, *The Equity Risk Premium: The Long-Run Future of the Stock Market* (New York: John Wiley & Sons).

Diamond, Peter A., 1999, What Stock Market Returns to Expect for the Future?, *Center for Retirement at Boston College number 2.*

Diamond, Peter A., 2001, What Stock Market Returns to Expect for the Future: An Update, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns* (Princeton: Princeton University Press).

Elton, Edwin J., and Martin J. Gruber, 1995, *Modern Portfolio Theory and Investment Analysis*, 5th ed. (New York: John Wiley & Sons).

Fama, Eugene F., and Kenneth R. French, 2002, The Equity Premium, *Journal of Finance,* 57, 637-659.

Fama, Eugene F., and Kenneth R French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance,* 47, 427-465.

Goetzmann, William N., Roger G. Ibbotson, and Liang Peng, 2001, A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability, *Journal of Financial Markets,* 4, 1-32.

Goss, Stephen C., 2001, Equity Yield Assumptions Used by the Office of the Chief Actuary, Social Security Administration, to Develop Estimates for Proposals with Trust

Fund and/or Individual Account Investments, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board.
http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Graham, John R., and Campbell R. Harvey, 2002, Expectations of Equity Risk Premia, Volatility and Asymmetry, Working paper, Duke University.

Grinold, Richard, and Kenneth Kroner, 2002, The Equity Risk Premium, *Barclays Global Investors' Investment Insights* 5.

Harris, Robert S., and Felicia C. Marston, 2001, The Market Risk Premium: Expectational Estimates Using Analysts' Forecasts, *Journal of Applied Finance,* 11, 6-16.

Harvey, Andrew, 1990, *The Econometric Analysis of Time Series*, 2nd ed. (Cambridge, MA: MIT Press).

Ibbotson Associates, 2003, *Stocks, Bonds, Bills, and Inflation Yearbook* (Ibbotson Associates, Chicago).

Ibbotson Associates, 2003, *Stocks, Bonds, Bills, and Inflation Yearbook*, Valuation Edition (Ibbotson Associates, Chicago).

Ibbotson, Roger G., and Peng Chen, 2003, Long-Run Stock Returns: Participating in the Real Economy, *Financial Analysts Journal,* 59, 88-98.

Kaplan, Paul D., and James D. Peterson, 1998, Full-Information Industry Betas, *Financial Management,* 27, 85-93.

LaChance, Marie-Eve, and Olivia S. Mitchell, 2003, Understanding Individual Account Guarantees, University of Michigan Retirement Research Center working paper WP 2003-035.

Leibowitz, Martin L., 2001, Equity Risk Premium Forum, TIAA-CREF, New York City, November.
http://www.aimrpubs.org/ap/issues/v2002n1/toc.html

Longstaff, Francis A., 2002, The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices, Working paper #9312, NBER, forthcoming in *Journal of Business.*

Mamaysky, Harry, 2002, A Model for Pricing Stocks and Bonds with Default Risk, Yale ICF Working Paper No. 02-13.

Malkiel, Burton G., 1999, *A Random Walk Down Wall Street* (New York: W. W. Norton & Company).

Mehra, Rajnish, 2002, Equity Premium Puzzle, in *Mastering Investments*, Ed. James Pickford (London: Financial Times-Prentice Hall).

Mehra, Rajnish, and Edward C. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics,* 15, 145-161.

Pratt, Shannon, P., 1998, *Cost of Capital Estimation and Applications*, John Wiley & Sons, New York.

Ralfe, John, Cliff Speed, and Jon Palin, 2003, Pensions and Capital Structure: Why hold equities in the pension fund?, Society of Actuaries Symposium, Vancouver, June.

Ross, Stephen A., 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory,* 13, 341-360.

Schwert, G. William, 1990, Indexes of United States Stock Prices from 1802 to 1987, *Journal of Business,* 63, 399-426.

Shiller, Robert, 1989, *Market Volatility* (Cambridge, MA: MIT Press).

Shoven, John B., 2001, What Are Reasonable Long-Run Rates of Return to Expect on Equities?, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Siegel, Jeremy J., 1999, The Shrinking Equity Premium, *Journal of Portfolio Management,* 26, 10-17.

Siegel, Jeremy J., 2002, *Stocks for the Long Run: The Definitive Guide to Financial Market Returns and Long-Term Investment Strategies*, 3rd ed. (New York: McGraw-Hill).

Social Security Advisory Board, 2002, Fiscal Year Annual Report http://www.ssab.gov/annualreport2002.pdf

Tversky, Amos, and Daniel Kahneman, 1991, Loss Aversion and Riskless Choice: A Reference Dependent Model, *Quarterly Journal of Economics,* 152, 1039-1061.

Weil, Philippe, 1989, The Equity Premium Puzzle and the Risk-free Rate Puzzle, *Journal of Monetary Economics,* 24, 401-421.

Welch, Ivo, 2000, Views of Financial Economists on the Equity Premium and on Professional Controversies, *Journal of Business,* 73, 501-537.

Welch, Ivo, 2001, The Equity Premium Consensus Forecast Revisited, Cowles Foundation discussion paper, No. 1325.

Wendt, Richard, 2002, Understanding Equity Risk Premium, *Risk and Rewards*, No. 38, Society of Actuaries, February.

Wilson, Jack W., and Charles P. Jones, 2002, An Analysis of the S&P 500 Index and Cowles's Extensions: Price Indexes and Stock Returns, 1870-1999, *Journal of Business*, 75:3, 505-533.

Appendix A
Ibbotson Market Data 1926-2002*

	Common Stocks	U. S. Treasury Bills	Arithmetic Short-Horizon
Year	Total Annual Returns	Total Annual Returns	Equity Risk Premia
1926	11.62%	3.27%	8.35%
1927	37.49%	3.12%	34.37%
1928	43.61%	3.56%	40.05%
1929	- 8.42%	4.75%	-13.17%
1930	-24.90%	2.41%	-27.31%
1931	-43.34%	1.07%	-44.41%
1932	- 8.19%	0.96%	- 9.15%
1933	53.99%	0.30%	53.69%
1934	- 1.44%	0.16%	- 1.60%
1935	47.67%	0.17%	47.50%
1936	33.92%	0.18%	33.74%
1937	-35.03%	0.31%	-35.34%
1938	31.12%	- 0.02%	31.14%
1939	- 0.41%	0.02%	- 0.43%
1940	- 9.78%	0.00%	- 9.78%
1941	-11.59%	0.06%	-11.65%
1942	20.34%	0.27%	20.07%
1943	25.90%	0.35%	25.55%
1944	19.75%	0.33%	19.42%
1945	36.44%	0.33%	36.11%
1946	- 8.07%	0.35%	- 8.42%
1947	5.71%	0.50%	5.21%
1948	5.50%	0.81%	4.69%
1949	18.79%	1.10%	17.69%
1950	31.71%	1.20%	30.51%
1951	24.02%	1.49%	22.53%
1952	18.37%	1.66%	16.71%
1953	- 0.99%	1.82%	- 2.81%
1954	52.62%	0.86%	51.76%
1955	31.56%	1.57%	29.99%
1956	6.56%	2.46%	4.10%

Appendix A
Ibbotson Market Data 1926-2002*

Year	Common Stocks Total Annual Returns	U. S. Treasury Bills Total Annual Returns	Arithmetic Short-Horizon Equity Risk Premia
1957	-10.78%	3.14%	-13.92%
1958	43.36%	1.54%	41.82%
1959	11.96%	2.95%	9.01%
1960	0.47%	2.66%	- 2.19%
1961	26.89%	2.13%	24.76%
1962	- 8.73%	2.73%	-11.46%
1963	22.80%	3.12%	19.68%
1964	16.48%	3.54%	12.94%
1965	12.45%	3.93%	8.52%
1966	-10.06%	4.76%	-14.82%
1967	23.98%	4.21%	19.77%
1968	11.06%	5.21%	5.85%
1969	- 8.50%	6.58%	-15.08%
1970	4.01%	6.52%	- 2.51%
1971	14.31%	4.39%	9.92%
1972	18.98%	3.84%	15.14%
1973	-14.66%	6.93%	-21.59%
1974	-26.47%	8.00%	-34.47%
1975	37.20%	5.80%	31.40%
1976	23.84%	5.08%	18.76%
1977	- 7.18%	5.12%	-12.30%
1978	6.56%	7.18%	- 0.62%
1979	18.44%	10.38%	8.06%
1980	32.42%	11.24%	21.18%
1981	- 4.91%	14.71%	-19.62%
1982	21.41%	10.54%	10.87%
1983	22.51%	8.80%	13.71%
1984	6.27%	9.85%	- 3.58%
1985	32.16%	7.72%	24.44%
1986	18.47%	6.16%	12.31%
1987	5.23%	5.47%	- 0.24%
1988	16.81%	6.35%	10.46%
1989	31.49%	8.37%	23.12%

Appendix A
Ibbotson Market Data 1926-2002*

Year	Common Stocks Total Annual Returns	U. S. Treasury Bills Total Annual Returns	Arithmetic Short-Horizon Equity Risk Premia
1990	- 3.17%	7.81%	-10.98%
1991	30.55%	5.60%	24.95%
1992	7.67%	3.51%	4.16%
1993	9.99%	2.90%	7.09%
1994	1.31%	3.90%	- 2.59%
1995	37.43%	5.60%	31.83%
1996	23.07%	5.21%	17.86%
1997	33.36%	5.26%	28.10%
1998	28.58%	4.86%	23.72%
1999	21.04%	4.68%	16.36%
2000	- 9.11%	5.89%	-15.00%
2001	-11.88%	3.83%	-15.71%
2002	-22.10%	1.65%	-23.75%
mean=	**12.20%**	**3.83%**	**8.37%**
Standard Dev=	**20.49%**	**3.15%**	**20.78%**

* *2003 SBBI Yearbook pages 38 and 39*

Appendix B

Source	Risk-free-Rate	ERP Estimate	Real risk-free rate	Nominal risk-free rate	Geometric	Arithmetic	Long-horizon	Short-horizon	Short-run expectation	Long-run expectation	Conditional	Unconditional
Historical												
Ibbotson Associates	3.8% [7]	8.4% [31]		X		X		X		X		X
Social Security												
Office of the Chief Actuary [1]	2.3%,3.0% [8]	4.7%,4.0% [32]	X		X		X			X		X
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	X		X	X	X	X		X	X	
Peter Diamond	2.2% [10]	<4.8% [34]	X		X		X			X	X	
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	X		X		X			X	X	
John Shoven [4]	3.0%,3.5% [12]	3.0% to 3.5% [36]	X		X		X			X	X	
Puzzle Research												
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	X		X		X			X	X	
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	X		X		X			X	X	
John Campbell and Robert Shiller	N/A	Negative [39]	X		?		?		X		X	
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]		X		X	X			X	X	
George Constantinides	2.0% [16]	6.9% [41]	X			X		X		X		X
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]		X		X	X	X		X	X	
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	X			X		X		X	X	
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	X			X		X		X		X
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]		X		X	X		X		X	
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	X		X	X	X			X		X
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	X		X		X			X	X	
Jeremy Siegel	3.5% [23]	2-3% [48]	X		X		X			?	X	
Surveys												
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]		X		?	X		X		X	
Ivo Welch	N/A [25]	7% [50]		X		X		X		X	X	
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]		X		X		X		X	X	
Misc.												
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]		X	X		X		X		X	
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]		X		X		X		X		X
Burton Malkiel	5.25% [29]	2.75% [54]		X	X		X			X	X	
Richard Wendt [6]	5.5% [30]	3.3% [55]		X		X	X			X	X	

Long-run expectation considered to be a forecast of more than 10 years.
Short-run expectation considered to be a forecast of 10 years or less.

Source	Risk-free Rate	ERP Estimate	Data Period	Methodology
Historical				
Ibbotson Associates	3.8% [7]	8.4% [31]	1926-2002	Historical
Social Security				
Office of the Chief Actuary [1]	2.3%, 3.0% [8]	4.7%, 4.0% [32]	1900-1995, Projecting out 75 years	Historical
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	Projecting out 75 years	Historical & Ratios (Div/Price & Earn Gr)
Peter Diamond	2.2% [10]	<4.8% [34]	Last 200 yrs for eq/ 75 for bonds, Proj 75 yrs	Fundamentals: Div Yld, GDP Gr
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	Projecting out 75 years	Fundamentals: Div/Price
John Shoven [4]	3.0%, 3.5% [12]	3.0% to 3.5% [36]	Projecting out 75 years	Fundamentals: P/E, GDP Gr
Puzzle Research				
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	1802 to 2001, normal	Fundamentals: Div Yld & Gr
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	Past 74 years, 74 year projection [56]	Fundamentals: Div Yld & Gr
John Campbell and Robert Shiller	N/A	Negative [39]	1871 to 2000, ten-year projection	Ratios: P/E and Div/Price
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]	1985-1998, long-term	Abnormal Earnings model
George Constantinides	2.0% [16]	6.9% [41]	1872 to 2000, long-term	Hist. and Fund.: Price/Div & P/E
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]	1926-1997, long run forward-looking	Weighing theoretical and empirical evid
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	1900-2000, prospective	Adj hist ret, Var of Gordon gr model
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	Estimate for 1951-2000, long-term	Fundamentals: Dividends and Earnings
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]	1982-1998, expectational	Fin analysts' est, div gr model
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	1926-2000, long-term	Historical and supply side approaches
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	1871 to 1998, forward-looking	Fundamentals: P/E, Div Yld, Div Gr
Jeremy Siegel	3.5% [23]	2-3% [48]	1802-2001, forward-looking	Earnings yield
Surveys				
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]	2Q 2000 thru 3Q 2002, 1 & 10 year proj	Survey of CFO's
Ivo Welch	N/A [25]	7% [50]	30-Year forecast, surveys in 97/98 & 99	Survey of financial economists
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]	30-Year forecast, survey around August 2001	Survey of financial economists
Misc.				
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]	Long-run (10-year) expected return	Fundamentals: Inc, Earn Gr, & Repricing
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]	1926-1997	Predominantly Historical
Burton Malkiel	5.25% [29]	2.75% [54]	1926 to 1997, estimate millennium [57]	Fundamentals: Div Yld, Earn Gr
Richard Wendt [6]	5.5% [30]	3.3% [55]	1960-2000, estimate for 2001-2015 period	Linear regression model

Footnotes:

[1] Social Security Administration.

[2] Presented to the Social Security Advisory Board.

[3] Presented to the Social Security Advisory Board. Update of 1999 article.

[4] Presented to the Social Security Advisory Board.

[5] Update to Welch 2000.

[6] Newsletter of the Investment Section of the Society of Actuaries.

[7] Arithmetic mean of U.S. Treasury bills annual total returns from 1926-2002.

[8] 2.3% Long-run real yield on Treasury bonds; used for Advisory Council proposals. 3.0% Long-term real yield on Treasury bonds; used in 1999 Social Security Trustees Report.

[9] Estimate for safe real interest rates in the future based on yield of long-term inflation-indexed Treasury securities of 3.5% and short-term real interest rates recently averaging about 3%.

[10] Real long-term bond yield using 75 year historical average.

[11] Real yield on long-term Treasuries (assumption by OCACT).

[12] 3.0% is the OCACT assumption. 3.5% is the real return on long-run (30-year) inflation-indexed Treasury securities.

[13] Long-term expected real geometric bond return (10 year-horizon).

[14] The yield on US government inflation-indexed bonds (starting bond real yield in Jan 2000).

[15] Average 10-year Government T-bond yield between 1985 and 1998 (yield of 11.43% in 1985 to 5.64% in 1998. The mean 30-year risk-free rate for each year of the U.S. sample period is 31 basis points higher than the mean 10-year risk-free rate.

[16] Rolled-over real arithmetic return of three-month Treasury bills and certificates.

[17] Historical 20-year Treasury bond return of 5.6%. Yield on 20-year Treasury bonds in 1998 was approximately 6%. Historical 1-month Treasury bill return of 3.8%. Yield on 1-month Treasury bills in 1998 was approximately 4%.

[18] United States historical arithmetic real Treasury bill return over 1900-2000 period. 0.9% geometric Treasury bill return.

[19] Average real return on six-month commercial paper (proxy for risk-free interest rate). Substituting the one-month Treasury bill rate for the six-month commercial paper rate causes estimates of the annual equity premium for 1951-2000 to rise by about 1.00%.

[20] Average yield to maturity on long-term U.S. government bonds, 1982-1998.

[21] Real, geometric risk-free rate. Geometric risk-free rate with inflation (nominal) 5.13%.
Nominal yield equivalent to historical geometric long-term government bond income return for 1926-2000.

[22] The ten- and thirty-year TIPS bond yielded 4.0% in August 1999.

[23] Return on inflation-indexed securities.

[24] Current 10-year Treasury bond yield. Survey administered from June 6, 2000 to June 4, 2002. The rate on the 10-year Treasury bond changes in each survey. For example, in the Dec. 1, 2000 survey, the current annual yield on the 10-year Treasury bond was 5.5%. For the June 6, 2001 survey, the current annual yield on the 10-year Treasury bond was 5.3%.

[25] Arithmetic per-annum average return on rolled-over 30-day T-bills.

[26] Average forecast of arithmetic risk-free rate of about 5% by deducting ERP from market return.

[27] Current nominal 10-year bond yield.

[28] Return on Treasury bills. Treasury bills yield of about 5 percent in mid-1998. Average historical return on Treasury bills 3.8 percent.

[29] Good quality corporate bonds will earn approximately 6.5% to 7%. Long-term zero-coupon Treasury bonds will earn about 5.25%. Long-term TIPS will earn a real return of 3.75%.

[30] 1/1/01 Long T-Bond yield; uses initial bond yields in predictive model.

[31] Arithmetic short-horizon expected equity risk premium. Arithmetic intermediate-horizon expected equity risk premium 7.4%. Arithmetic long-horizon expected equity risk premium 7.0%. Geometric short-horizon expected equity risk premium 6.4%.

[32] Geometric equity premium over long-term Treasury securities. OCACT assumes a constant geometric real 7.0% stock return.

[33] Long-run average equity premium of 1.5% to 2.5% in geometric terms and 3% to 4% in arithmetic terms.

[34] Lower return over the next decade, followed by a geometric, real 7.0% stock return for remaining 65 years or lower rate of return for entire 75-year period (obscures pattern of returns).

[35] Most likely poor return over the next decade followed by a return to historic yields. Working from OCACT stock return assumption, he gives a single rate of return on equities for projection purposes of 6.0 to 6.5% (geometric, real).

[36] Geometric real stock return over the geometric real return on long-term government bonds.

[37] Expected geometric return over long-term government bonds. Their current risk premium is approximately zero, and their recommended expectation for the future real return for both stocks and bonds is 2-4 percent. The "normal" level of the risk premium is modest (2.4 percent or quite possibly less).

[38] Geometric real returns on stocks are likely to be in the 3%-4% range for the foreseeable future (10-20 years).

[39] Substantial declines in real stock prices, and real stock returns below zero, over the next ten years (2001-2010).

[40] The equity premium for each year between 1985 and 1998 in the United States. Similar results for five other markets.

[41] Unconditional, arithmetic mean aggregate equity premium over the 1872-2000 period. Over the period 1951 to 2000, the adjusted estimate of the unconditional mean premium is 6.0%. The corresponding estimate over the 1926 to 2000 period is 8.0%. Sharp distinction between conditional, short-term forecasts of the mean equity return and premium and estimates of the unconditional mean.

[42] Long run arithmetic future ERP of 3.5% to 5.5% over Treasury bonds and 5% to 7% over Treasury bills. Compares estimates to historical returns of 7.4% for bond premium and 9.2% for bill premium.

[43] 5.4% United States arithmetic expected future ERP relative to bills. 4.0% World (16 countries) arithmetic expected future ERP relative to bills. 4.1% United States geometric expected future ERP relative to bills. 3.0% World (16 countries) geometric expected future ERP relative to bills.

[44] 3.83% unconditional expected annual simple equity premium return (referred to as the annual-bias adjusted estimate of the annual equity premium) using dividend growth model. 4.78% unconditional expected annual simple equity premium return (referred to as the annual-bias adjusted estimate of the annual equity premium) using earnings growth model. Compares these results against historical real equity risk premium of 7.43% for 1951-2000.

[45] Average expectational risk premium. Because of the possible bias of analysts' optimism, the estimates are interpreted as "upper bounds" for the market premium. The average expectational risk premium is approximately equal to the arithmetic (7.5%) long-term differential between returns on stocks and long-term government bonds.

[46] 4% geometric (real) and 6% arithmetic (real). Forward looking long-horizon sustainable equity risk premium.

[47] Using the dividend discount model, the forward-looking real long-term geometric return on equity is 3.3%. Based on the earnings yield, the forward-looking real long-term geometric retum on equity is between 3.1% and 3.7%.

[48] Future geometric equity premium. Future real return on equities of about 6%.

[49] The 10-year premium. The one-year risk premium averages between 0.4 and 5.2% depending on the quarter surveyed.

[50] Arithmetic 30-year forecast relative to short-term bills; 10-year same estimate. Second survey 6.8% for 30 and 10-year estimate. 1-year horizon between 0.5% and 1.5% lower. Geometric 30-year forecast around 5.2% (50% responded to this question).

[51] Arithmetic 30-year equity premium (relative to short-term T-bills). Geometric about 50 basis points below arithmetic. Arithmetic 1-year equity premium 3 to 3.5%.

[52] 2.5% current (conditional) geometric equity risk premium. 3.25% long-run, geometric normal or equilibrium equity risk premium.

[53] Extra arithmetic return versus Treasury bills. "Brealey and Myers have no official position on the exact market risk premium, but we believe a range of 6 to 8.5 percent is reasonable for the United States. We are most comfortable with figures towards the upper end of the range."

[54] The projected geometric (nominal) total return for the S&P 500 is 8 percent per year.

[55] Arithmetic mean 15 year horizon.

[56] 74 years since Dec 1925 and 74 years starting Jan 2000.

[57] Estimate the early decades of the twenty-first century.

Appendix C
Estimating a Short-Horizon Arithmetic Unconditional Equity Risk Premium

Source	Risk-free Rate	ERP Estimate	Stock Return Estimate	Geometric to arithmetic	Real to nominal	Conditional to unconditional [60]	Fixed short-horizon RFR	Short-horizon arithmetic unconditional ERP estimate
	I	II	III	IV	V	VI	VII	VIII
Historical								
Ibbotson Associates	3.8% [7]	8.4% [31]	12.2%	0.0%	0.0%	0.00%	3.8%	8.4%
Social Security								
Office of the Chief Actuary [1]	2.3%,3.0% [8]	4.7%,4.0% [32]	7.0%	2.0%	3.1%	0.00%	3.8%	8.3%
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	6.0%-7.5%	0.0%	3.1%	0.46%	3.8%	5.8%-7.3%
Peter Diamond	2.2% [10]	<4.8% [34]	<7.0%	2.0%	3.1%	0.46%	3.8%	<8.8%
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	6.0%-6.5%	2.0%	3.1%	0.46%	3.8%	7.8%-8.3%
John Shoven [4]	3.0%,3.5% [12]	3.0% to 3.5% [36]	6.0%-7.0%	2.0%	3.1%	0.46%	3.8%	7.8%-8.8%
Puzzle Research								
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	6.1%	2.0%	3.1%	0.46%	3.8%	7.9%
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	3.2%	2.0%	3.1%	0.46%	3.8%	5.0%
John Campbell and Robert Shiller	N/A	Negative [39]	Negative	N/A	N/A	N/A	N/A	N/A
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]	11.03%	0.0%	0.0%	0.46%	3.8%	7.69%
George Constantinides	2.0% [16]	6.9% [41]	8.9%	0.0%	3.1%	0.00%	3.8%	8.2%
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]	8.8%-10.8%	0.0%	0.0%	0.46%	3.8%	5.5%-7.5%
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	6.4% [58]	0.0%	3.1%	0.46%	3.8%	6.2% [61]
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	7.07%-8.02%	0.0%	3.1%	0.00%	3.8%	6.37%-7.32%
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]	12.34% [59]	0.0%	0.0%	0.46%	3.8%	9.00%
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	8.05%	0.0%	3.1%	0.00%	3.8%	7.35%
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	3.1%-3.7%	2.0%	3.1%	0.46%	3.8%	4.9%-5.5%
Jeremy Siegel	3.5% [23]	2-3% [48]	5.5%-6.5%	2.0%	3.1%	0.46%	3.8%	7.3%-8.3%
Surveys								
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]	8.3%-10.2%	N/A	0.0%	0.46%	3.8%	5.0%-6.9%
Ivo Welch	N/A [25]	7% [50]	N/A	0.0%	0.0%	0.46%	0.0%	7.5%
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]	10.0%-10.5%	0.0%	0.0%	0.46%	3.8%	6.7%-7.2%
Misc.								
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]	7.5%,8.25%	2.0%	0.0%	0.46%	3.8%	6.16%-6.91%
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]	N/A	0.0%	0.0%	0.00%	0.0%	6.0%-8.5%
Burton Malkiel	5.25% [29]	2.75% [54]	8.0%	2.0%	0.0%	0.46%	3.8%	6.7%
Richard Wendt [6]	5.5% [30]	3.3% [55]	8.8%	0.0%	0.0%	0.46%	3.8%	5.5%

Column formulas:
III = I + II
VIII = III + IV + V + VI – VII

Source for adjustments:
2003 Ibbotson Yearbook Table 2-1 page 33
Fama French 2002 (see footnote 60)

Footnotes (1-57 from Appendix B):
[58] World estimate of 5.0%.
[59] Long risk-free of 5.2% plus 7.14%.
[60] For the 1951-2000 period, Fama and French (2002) adjust the conditional dividend growth model estimate upwards by 1.28% for an unconditional estimate, and they make a 0.46% upwards adjustment to the earnings growth model. We select the smaller of the two as an approximate minimum adjustment. For the longer period of 1872-2000, a comparable adjustment would be 0.82% for the dividend growth model and 0.54% for the 1872-1950 period using a dividend growth model. Earnings growth rates are shown by Fama and French only for the 1951-2000 period.
[61] World estimate of 4.8%.

Appendix D

Historical and Forecasted Equity Returns- All Ibbotson and Chen Models (Percent).

Method/ Model	Sum	Inflation	Real Risk-Free Rate	Equity Risk Premium	Real Capital Gain	g(Real EPS)	g(Real Div)	- g (Pay out Ratio)	g (BV)	g (ROE)	g P/E)	g(Real GDP/ POP)	g(FS-GDP/ POP)	Income Return	Re-Investment + Interaction	Additional Growth	Forecast Earnings Growth
Column #	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
Historical																	
Method 1	10.70	3.08	2.05	5.24											0.33		
Method 2	10.70	3.08			3.02									4.28	0.32		
Method 3	10.70	3.08				1.75					1.25			4.28	0.34		
Method 4	10.70	3.08					1.23	0.51			1.25			4.28	0.35		
Method 5	10.70	3.08							1.46	0.31	1.25			4.28	0.31		
Method 6	10.70	3.08										2.04	0.96	4.28	0.32		
Forecast with Historical Dividend Yield																	
Model 3F	9.37	3.08				1.75								4.28	0.26		
Model 3F (ERP)	9.37	3.08	2.05	3.97											0.27		
Forecast with Current Dividend Yield																	
Model 4F	5.44	3.08					1.23							1.10 [a]	0.03		
Model 4F (ERP)	5.44	3.08	2.05	0.24											0.07		
Model 4F_2	9.37	3.08					1.23	0.51						2.05 [b]	0.21	2.28	
Model 4F_2 (FG)	9.37	3.08												1.10 [a]	0.21		4.98

Source: The data and format was made available by Ibbotson/Chen and is reprinted with permission by Ibbotson Associates.
Corresponds to Ibbotson/Chen Table 2 Exhibit; column numbers have been added.
[a] 2000 dividend yield
[b] Assuming the historical average dividend-payout ratio, the 2000 dividend yield is adjusted up 0.95 pps.

	Formula*	Description of Method
Historical		
Method 1	I=(1+II)*(1+III)*(1+IV)-1	Building Blocks Method: inflation, real risk-free rate, and equity risk premium.
Method 2	I=[(1+II)*(1+V)-1]+XIV+XV	Capital Gain and Income Method: inflation, real capital gain, and income return.
Method 3	I=[(1+II)*(1+VI)*(1+XI)-1]+XIV+XV	Earnings Model: inflation, growth in earnings per share, growth in price to earnings ratio, and income return.
Method 4	I=[(1+II)*(1+XI)*(1+VII)/(1-VIII)-1]+XIV+XV	Dividends Model: inflation, growth rate of price earnings ratio, growth rate of the dollar amount of dividends after inflation, growth rate of payout ratio, and dividend yield (income return).
Method 5	I=[(1+II)*(1+XI)*(1+IX)*(1+X)-1]+XIV+XV	Return on Book Equity Model: inflation, growth rate of price earnings ratio, growth rate of book value, growth rate of ROE, and income return.
Method 6	I=[(1+II)*(1+XII)*(1+XIII)-1]+XIV+XV	GDP Per Capita Model: inflation, real growth rate of the overall economic productivity (GDP per capita), increase of the equity market relative to the overall economic productivity, and income return.
Forecast with Historical Dividend Yield		
Model 3F	I=[(1+II)*(1+VI)-1]+XIV+XV	Forward-looking Earnings Model: inflation, growth in real earnings per share, and income return.
Model 3F (ERP)	IV=(1+I)/[(1+II)*(1+III)]-1	Using Model 3F result to calculate ERP.
Forecast with Current Dividend Yield		
Model 4F	I=[(1+II)*(1+VII)-1]+XIV+XV	Forward-looking Dividends Model: inflation, growth in real dividend, and dividend yield (income return); also referred to as Gordon model.
Model 4F (ERP)	IV=(1+I)/[(1+II)*(1+III)]-1	Using Model 4F result to calculate ERP.
Model $4F_2$	I=[(1+II)*(1+VII)*(1+VIII)-1]+XIV+XV+XVI	Attempt to reconcile Model 4F and Model 3F.
Model $4F_2$ (FG)	XVII=[(1+I)/(1+II)-1]-XIV-XV	Using Method $4F_2$ result to calculate forecasted earnings.

Explanation of Ibbotson/Chen Table 2 Exhibit; using column numbers to represent formula.

Gradients of Risk Measures: Theory and Application to Catastrophe Risk Management and Reinsurance Pricing

John A. Major, ASA, MAAA

Gradients of Risk Measures: Theory and Application to Catastrophe Risk Management and Reinsurance Pricing

By John A. Major, ASA[1]

As part of pricing, many reinsurers would like to know the incremental impact that adding a new contract or canceling an existing contract might have on the capital needed to support the entire portfolio of business. Typically, catastrophe models take hours or days to run, ruling out a straightforward approach. A method of assessing incremental impact which did not require repeatedly simulating losses to the entire portfolio would therefore be quite useful. Efficient procedures for calculating the first derivatives of widely-used risk measures (such as Value at Risk and Tail Value at Risk) with respect to portfolio parameters would support the development of such a method. This paper presents general formulas for gradients of risk measures including VaR and TVaR. While the derivative of VaR in the case of linear risk models is widely known, this paper presents the general solution applicable not only to linear portfolio weights, but also to nonlinear parameters, such as retentions and limits. Implementation of the theoretical formulas within existing catastrophe simulation models is elaborated. A normal mixture approximation leads to a closed form solution for the incremental impact on VaR or TVaR of adding or removing a contract from a portfolio of excess-of-loss contracts.

1. Introduction

Reinsurance pricing is, in part, a portfolio problem: a matter of determining a sufficiently high expected return (margin in the premium) to compensate the firm for the risk it is taking on. In a survey of the practice of catastrophe excess-of-loss (XOL) pricing among reinsurers, Major and Kreps [2002] find the more sophisticated markets "attempt to assess the incremental impact a new contract would have on the capital needed to support the overall portfolio...." (Models of such an approach can be seen in Kreps [1990; 1999]. Tasche [1999] centers a theory of capital allocation and portfolio component returns on the first derivative of the risk measure.) They also state that most markets "lacked the analytical skills or the brute force computing power needed to accomplish it." Typically, catastrophe models running on the entire portfolio take hours or days so runs are done only a few times per year, ruling out incremental analysis of new contract proposals. Therefore a method of assessing incremental impact which did not require simulating losses to the entire portfolio would be quite useful.

Key statistics for measuring risk in the context of insurance capital requirements include Probable Maximum Loss, also known as Value at Risk or VaR, and Expected Shortfall, which is closely related to Tail Value at Risk or TVaR. Efficient procedures for calculating the first derivatives of these risk measures with respect to portfolio parameters would support the development of desired pricing algorithms. These risk measures also extend far beyond the domain of catastrophe insurance. Because of the Basel Committee [1996], VaR is used heavily by nearly all banks and financial institutions. TVaR is also becoming of considerable interest to financial risk managers (Andersson et. al. [2000], Gaivoronski and Pflug [2000], Yamai and Yoshiba [2002]). Efficient computation of gradients would then support efforts in risk management throughout the financial services industry.

This paper presents general formulas for the gradients of risk measures including VaR and TVaR with respect to parameters of a loss function (which may be nonlinear) over an arbitrary smooth distribution of risk factors. While the derivative of VaR in the case of linear risk models is widely known (see, for example, Gourieroux et. al. [2000]) a theorem of Uryasev [1995a,b; 1999] is applied to derive the general solution applicable to nonlinear parameters, such as retentions and limits, as well as to linear portfolio weights.

[1] email: john.a.major@guycarp.com

Specializations to linear and quadratic loss functions and analytic solutions for normally distributed risk factors will be discussed. A normal mixture approximation will be shown to lead to a closed form solution for the incremental impact on VaR or TVaR of adding or removing a contract from a portfolio of piecewise linear excess-of-loss contracts.

Section 2 of the paper provides the theoretical results. First the risk measures of interest are defined, and the role of the gradient in first-order Taylor series expansions and optimization is outlined. Second, the main theorems representing gradients as functions of expectations are given. Section 3 presents the bilinear special case and applications to insurance and reinsurance problems.

Section 4 discusses the efficient implementation of the theoretical formulas within existing Monte Carlo catastrophe simulation models. Section 5 addresses analytical solutions and approximations, including the delta-normal and gamma-normal models as well as conditional normal models. Ultimately the piecewise-linear normal-mixture is revealed as a generic analytic approximation. Section 6 showcases the results of a simulation study comparing "brute force" and theorem-based Monte Carlo methods to the re/insurance applications.

Section 7 provides an overview of some related work. Section 8 concludes. Section 9 is an appendix with a dictionary of mathematical symbols and selected derivations.

2. Theory

2.1 Mathematical framework

Let **X** in R^N represent a random variable that underlies a loss process (think of hurricane wind speeds or earthquake peak ground acceleration at various locations). Let $f(\mathbf{x})$ be the (multivariate) probability density function. The pdf f is assumed smooth for the theoretical development, but this will be relaxed in the discussion of Monte Carlo simulation later. **X** describes the "hazard" independently of any particular portfolio of exposures.[2] The components of **X** are often referred to as "risk factors."

Let $L = g(\mathbf{X},\theta)$ define the net losses to a portfolio of insured exposures, where θ in R^M is a parameter describing the exposure (think of all the financial specifications of a reinsurance deal: exposures at various locations, as well as retentions, limits, sublimits, etc.).

The cumulative probability distribution of net losses is given by the multidimensional integral

$$F(l;\theta) = \Pr\{L \le l\} = \int_{\{x:g(x,\theta)\le l\}} f(\mathbf{x})dV$$

where $dV = dx_1dx_2\ldots dx_N$ represents a volume integral element. That is, integrating (or summing) all the probability for events which cause losses less than or equal to l gives the cumulative probability. Note that the set $\{\mathbf{x}|g(\mathbf{x},\theta) = l\}$ is an (N-1)-dimensional submanifold of R^N which divides the space of **X** points into two regions: those that result in losses less than l and those that result in losses greater than l. Call this boundary surface an *isoclast.*[3]

---

[2] The reader may regard this discussion as being concerned with severity or per-occurrence distributions only. For an extension to compound processes, see Daykin et. al. [1994].

[3] From the Greek "*isos*" = "equal" and "*klastos*" = "broken."

2.2 Defining risk measures

The cumulative probability distribution of L can be cumbersome to deal with. It is a function and requires an infinite set (theoretically) or at least a very large set (practically) of numbers to describe it. One often seeks scalar measures of risk to characterize the distribution more economically.

2.2.1 Expectations and moments

One class of risk measures consists of mathematical expectations of functions of the random loss E[h(L)]. The simplest nontrivial form of this is the first moment, or mean:

$$\mu[L] = E[L] = \int g(\mathbf{x};\boldsymbol{\theta}) \cdot f(\mathbf{x})dV$$

The mean is a measure of central tendency or location. In statistical practice, it is usually associated with a companion measure of spread, the second moment (variance) or its square root, the standard deviation:

$$\sigma^2[L] = E\left[(L-\mu)^2\right] = \int \left(g(\mathbf{x};\boldsymbol{\theta}) - \mu\right)^2 \cdot f(\mathbf{x})dV$$

$$\sigma[L] = \sqrt{\sigma^2[L]}$$

2.2.2 Exceedance probabilities

While it is inconvenient to assess the entire distribution, it may be valuable to capture certain points on the curve. For example, the loss of particular dollar amounts corresponding to various levels of financial distress (e.g., insolvency or regulatory thresholds) may be of interest. Define the exceedance probability at loss amount l as:

$$Q(l;\boldsymbol{\theta}) = \Pr\{L > l\} = 1 - F(l) = \int_{g(\mathbf{x};\boldsymbol{\theta})>l} f(\mathbf{x})dV$$

2.2.3 VaR L_q, and TVaR T_q

Value at Risk (VaR) answers the question: "If this portfolio experiences an adverse outcome, how bad could it be?" An exercise in measuring VaR consists of specifying the time and probability level and quantifying the monetary outcome (Jorion [2000]). Because time and probability must be specified, there are many VaR measurements that can be taken on a particular portfolio. In this paper, it is assumed that one time horizon (not symbolized) and one probability level (symbolized by q) have been specified. In application, many measurements can (and should) be taken for a more complete understanding of the risk.

Define the Value-at-Risk (VaR) $\mathrm{L}_q(\boldsymbol{\theta})$ as the solution to the equation

$$Q\left(\mathrm{L}_q(\boldsymbol{\theta});\boldsymbol{\theta}\right) = q.$$

That is, VaR is the level of losses that will only be exceeded with probability q. Note that it is a function of the portfolio description $\boldsymbol{\theta}$.

The typical VaR application in finance has **X** representing changes to current values of key interest rates, foreign exchange rates, etc., and L the resulting change in value of an investment portfolio. The parameter $\boldsymbol{\theta}$ might be implicit in the computation of g, especially if the task is to compute VaR on the current portfolio, without regard to changes in the portfolio composition. When explicit, θ may represent quantities of various securities or classes of securities held.

A pure insurance species of VaR is Probable Maximum Loss (PML). Here, **X** represents the random contingencies affecting a portfolio of insurance liabilities and L the associated loss. The canonical example from property insurance is the use of a catastrophe model to answer the question "What catastrophe loss will not be exceeded, on average, once in a hundred years?" The "100-year return period loss" is the one-year VaR corresponding to $q = 0.01$.[4] Similar questions about net operating results, including underwriting gain as well as investment income, are now routine in insurance applications of Dynamic Financial Analysis (DFA).

Tail Value at Risk (TVaR) is defined as the conditional expectation of the loss (the average loss) given that the loss is greater than the VaR. That is, it is the average loss among only those losses that exceed the VaR:

$$\mathsf{T}_q(\theta) = \frac{1}{q} \cdot \int_{\{\mathbf{x}:g(\mathbf{x},\theta) > \mathsf{L}_q(\theta)\}} g(\mathbf{x},\theta) \cdot f(\mathbf{x}) dV$$

TVaR is gaining favor relative to VaR because it possesses a property known as *coherence* (Artzner et. al. [1999]).

2.2.4 The gradient operator ∇

Managing a portfolio means observing and controlling (or at least influencing) its changes over time to achieve desired outcomes. The gradient of a risk measure – formally, the vector of partial derivatives – with respect to changes in parameters defining the portfolio provides a "sensitivity map" that shows explicitly how any small change to the portfolio will change the measure. Coupled with other sensitivity measures, the gradient can be used to design a change strategy aimed at improving the risk/reward characteristics of the portfolio.

If a portfolio described by θ has a risk measure $\mathsf{R}(\theta)$, and the portfolio parameter θ is perturbed by a small $\Delta\theta$, then the new value of R can be approximated by a first-order Taylor expansion:

$$\mathsf{R}(\theta + \Delta\theta) = \mathsf{R}(\theta) + \nabla_\theta \mathsf{R} \cdot \Delta\theta$$

where, in the case of an (M=1)-dimensional parameter θ, $\nabla_\theta \mathsf{R}$ is the first derivative of $\mathsf{R}(\theta)$ with respect to θ, or, in the general vector case (M>1), the vector of partial derivatives, i.e., the *gradient vector* of $\mathsf{R}(\theta)$ with respect to the portfolio vector parameter. Specifically,

$$\nabla_\theta \mathsf{R} = \left\langle \frac{\partial \mathsf{R}}{\partial \theta_1}, \frac{\partial \mathsf{R}}{\partial \theta_2}, \frac{\partial \mathsf{R}}{\partial \theta_3}, \ldots, \frac{\partial \mathsf{R}}{\partial \theta_M} \right\rangle$$

Writing out the vector multiplication term-by term, it would look like:

---

[4] Actually, the 100-year return period loss corresponds to VaR at q=0.00995 under Poisson frequency assumptions, but this distinction is often ignored in practice.

$$\nabla_\theta R \cdot \Delta\theta = \sum_j \left(\frac{\partial R(\theta)}{\partial \theta_j} \cdot \Delta\theta_j \right)$$

where the j indexes over all the components of θ.

The significance of the gradient is this: with the gradient in hand, the incremental impact of proposed (small) changes to the portfolio can be evaluated *without recomputing R*. The gradient can be used as part of an optimization procedure to seek out "change directions" for portfolio improvement, thereby reducing the number of computations of R in the search. Without using the gradient, if the dimension M of the portfolio parameter is large, it may be impractical to schedule all the required recomputations of R, even if a single cycle is fairly efficient. Additionally, if the model is analytically tractable, optimal change vectors might be obtained by a closed-form solution. Both of these points will be elaborated below.

2.3 Main theorems – gradients as integrals and expectations

2.3.1 Gradient of moments

The gradients of moments follow from smoothness giving us the ability to interchange differentiation and integration:

$$\nabla_\theta \mu[L] = \nabla_\theta \int g(\mathbf{x};\theta) \cdot f(\mathbf{x})dV = \int \nabla_\theta g(\mathbf{x};\theta) \cdot f(\mathbf{x})dV = E[\nabla_\theta g(\mathbf{X},\theta)]$$

where the subscript θ on the gradient symbol ∇ is there to emphasize that the partial derivatives are with respect to the θ arguments, not the $\mathbf{x}$ arguments.[5] This equation, as written, involves the integral of vectors, but it can be interpreted one term at a time by substituting $\partial/\partial\theta_j$ for the ∇ symbol.

The reader can verify that

$$\nabla_\theta \sigma^2[L] = 2 \cdot \int g \cdot \nabla_\theta g \cdot f(\mathbf{x})dV - \mu \cdot \nabla_\theta \mu = 2 \cdot E[(g-\mu) \cdot \nabla_\theta g(\mathbf{X},\theta)]$$

$$\nabla_\theta \sigma[L] = \frac{1}{2 \cdot \sigma} \cdot \nabla_\theta \sigma^2[L] = E\left[\frac{g-\mu}{\sigma} \cdot \nabla_\theta g(\mathbf{X},\theta)\right]$$

2.3.2 "Integral over the surface formula"

The most direct route to deriving the gradients of exceedance probabilities is to make use of the following result (Uryasev [1995a, b; 1999]). Let the domain of integration be defined by $\rho(\psi) = \{ \mathbf{x} \in R^N \mid \gamma(\mathbf{x},\psi) \le 0 \}$ and by $\partial\rho$ the boundary of this set. Consider the volume integral $H(\psi) = \int_{\rho(\psi)} \phi(\mathbf{x},\psi)dV$ where dV denotes $\mathbf{x}$-integration as before. Note that both the integrand and the region of integration are functions of a free parameter ψ. If the constraint function γ is differentiable and the following integrals exist, then the gradient of H with respect to ψ is given by:

---

[5] Superficially, it appears that Stokes' Theorem – $\int_\rho \nabla_\mathbf{x} g dV = \int_{\partial\rho} g dS$ – might be relevant here, but we are dealing with ∇_θ, not $\nabla_\mathbf{x}$.

$$\nabla_\psi H(\psi) = \int_{\rho(\psi)} \nabla_\psi \phi(\mathbf{x},\psi) dV - \int_{\partial\rho(\psi)} \frac{\nabla_\psi \gamma(\mathbf{x},\psi)}{\|\nabla_\mathbf{x} \gamma(\mathbf{x},\psi)\|} \cdot \phi(\mathbf{x},\psi) dS$$

where dS denotes (hyper)surface element and $\|\mathbf{v}\|$ denotes vector norm.[6]

2.3.3 Gradients of exceedance probabilities

The derivative of $Q(L,\theta)$ with respect to L can be obtained by taking $\psi = L$, $H(\psi) = Q(L,\theta)$, $\phi(\mathbf{x},\psi) = f(\mathbf{x})$, and $\gamma(\mathbf{x},\psi) = L - g(\mathbf{x},\theta)$, where θ is treated as a constant. Since ϕ is not a function of ψ, the first integral vanishes and one is left with:

$$\frac{\partial Q(L,\theta)}{\partial L} = -\int_{g(\mathbf{x},\theta)=L} \frac{1}{\|\nabla_\mathbf{x} g(\mathbf{x},\theta)\|} \cdot f(\mathbf{x}) dS = -C_N \cdot E\left[\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\|^{-1} \middle| g(\mathbf{X},\theta) = L\right]$$ where the prefactor $C_N = \int_{g(\mathbf{x};\theta)=L} f(\mathbf{x}) dS$.

The gradient of $Q(L,\theta)$ with respect to θ is obtained similarly by taking $\psi = \theta$, $H(\psi) = Q(L,\theta)$, $\phi(\mathbf{x},\psi) = f(\mathbf{x})$, and $\gamma(\mathbf{x},\psi) = L - g(\mathbf{x},\theta)$, where now L is treated as a constant:

$$\nabla_\theta Q(L,\theta) = -\int_{g(\mathbf{x},\theta)=L} \frac{-\nabla_\theta g(\mathbf{x},\theta)}{\|\nabla_\mathbf{x} g(\mathbf{x},\theta)\|} \cdot f(\mathbf{x}) dS = C_N \cdot E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\|} \middle| g(\mathbf{X},\theta) = L\right]$$

The relevance of $\|\nabla_\mathbf{x} g\|$, and how to avoid the necessity for computing it, is discussed in the section on Monte Carlo implementation.

2.3.4 Gradient of Value at Risk

Applying the implicit function theorem to the equation $Q(L_q(\theta),\theta) - q = 0$:

$$\frac{\partial L_q}{\partial \theta} = -\frac{\partial Q}{\partial \theta} \bigg/ \frac{\partial Q}{\partial L}.$$

Substituting the previously-obtained derivatives of Q:

$$\frac{\partial L_q}{\partial \theta} = -\frac{\partial Q/\partial \theta}{\partial Q/\partial L} = \frac{\int_{g(\mathbf{x};\theta)=L_q} \frac{\partial g}{\partial \theta} \|\nabla_\mathbf{x} g\|^{-1} \cdot f(\mathbf{x}) dS}{\int_{g(\mathbf{x};\theta)=L_q} \|\nabla_\mathbf{x} g\|^{-1} \cdot f(\mathbf{x}) dS} = \frac{E\left[\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\|^{-1} \cdot \nabla_\theta g(\mathbf{X},\theta) \mid g(\mathbf{X},\theta) = L_q\right]}{E\left[\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\|^{-1} \mid g(\mathbf{X},\theta) = L_q\right]}$$

---

[6] See previous footnote.

Corollary: Define the probability density function $h(\mathbf{x})$ on the $g(\mathbf{X};\theta) = \mathsf{L}_q$ isoclastic hypersurface by

$$h(\mathbf{x}) = \frac{\left\|\nabla_{\mathbf{x}} g\right\|^{-1} \cdot f(\mathbf{x})}{\int\limits_{g(\mathbf{x};\theta)=\mathsf{L}_q} \left\|\nabla_{\mathbf{x}} g\right\|^{-1} \cdot f(\mathbf{x}) dS}.$$ Then

$$\nabla_{\theta} \mathsf{L}_q = \mathsf{E}_h\left[\nabla_{\theta} g(\mathbf{X},\theta)\right]$$

where the expectation is taken with respect to the density $h(X)$, on its domain (the isoclast).

2.3.5 Gradient of Tail Value at Risk

With the result for the gradient of VaR in hand, the gradient of TVaR is obtained by a more-or-less straightforward application of the integral over the surface formula:

$$\nabla \mathsf{T}_q(\theta) = \frac{1}{q} \cdot \int_{\{\mathbf{x}:g(\mathbf{x},\theta)>\mathsf{L}_q(\theta)\}} \nabla_{\theta} g(\mathbf{x},\theta) \cdot f(\mathbf{x}) dV = E\left[\nabla_{\theta} g(\mathbf{X},\theta) \mid g(\mathbf{X},\theta) > \mathsf{L}_q(\theta)\right] = E_k\left[\nabla_{\theta} g(\mathbf{X},\theta)\right]$$

where $k(\mathbf{x})$ is the conditional distribution of $\mathbf{X}$ in the tail where $L > \mathsf{L}_q$. This is proved in the appendix (section 9.2).

Note the conceptual similarity between the gradient of VaR and the gradient of TVaR. Both are conditional expectations of $\nabla_{\theta} g$; they differ in the underlying conditional distribution on $\mathbf{X}$. For VaR, it is h, the distribution restricted to the isoclast (but reweighted for "thickness"). For TVaR, it is k, the distribution restricted to the tail, the half space where $\mathbf{X}$ puts losses over the VaR.

3. Applications

3.1 Bilinear, delta model

Perhaps the simplest nontrivial form is the bilinear loss function $L = g(\mathbf{X};\theta) = \theta^T \cdot \mathbf{X}$, where the superscript "T" indicates matrix transpose. The bilinear case occurs in property catastrophe loss modeling where elements of θ represent exposed property values at subscripted locations and the elements of $\mathbf{X}$ are corresponding damage rates (in a hurricane, say). L represents "ground-up" loss in CAT modeling parlance. This will be illustrated in the case study later. The bilinear case is by far the most frequently used form in financial modeling. There, $\mathbf{X}$ represents prices or returns of various asset classes and θ represents the quantity of each asset in a portfolio.

3.1.1 Moments

Calculation of moments is nearly trivial.

$\mu[L] = \theta^T \cdot \mu[\mathbf{X}]$ and $\sigma^2[L] = \theta^T \cdot vcv[\mathbf{X}] \cdot \theta$ where $vcv[\mathbf{X}] = E\left[(\mathbf{X}-\mu)\cdot(\mathbf{X}-\mu)^T\right]$, making

$\nabla_{\theta}\mu[L] = \mu[\mathbf{X}]$ and $\nabla_{\theta}\sigma^2[L] = 2 \cdot vcv[\mathbf{X}] \cdot \theta$ and $\nabla_{\theta}\sigma[L] = \sigma \cdot \theta$

3.1.2 Exceedance probabilities

Recalling the constant C_N from section 2.3.3, the simple form of g results in simple gradients:

$$\partial Q/\partial L = -C_N/\|\theta\| \text{ and } \nabla_\theta Q = (C_N/\|\theta\|)\cdot E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} = L].$$

3.1.3 Value at Risk and Tail Value at Risk

VaR and TVaR follow as: $\nabla_\theta L_q = E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} = L_q]$ and $\nabla_\theta T_q = E[\mathbf{X} \mid \theta^T \cdot \mathbf{X} > L_q]$. Note in the case of VaR that the "transformed distribution" h is simply the conditional distribution of $\mathbf{X}$ on the isoclast.

The gradient of VaR in the special bilinear case has been proved in many papers. See, e.g., Gourieroux et. al. [2000] and Tasche [1999].

3.1.4 Delta approximation

The bilinear case corresponds to the use of the "delta" approximation and most often emerges as part of the "delta-normal" model of VaR. The delta approximation writes:

$$L = g(\mathbf{x}_0 + \mathbf{X}) \approx g_0 + (\nabla_\mathbf{X} g)\cdot \mathbf{X}$$

and considers the vector $\delta = (\nabla_\mathbf{X} g)^T$ to comprise the portfolio parameter θ. The delta-normal model will be considered in section 5.1 in the context of analytic approximations.

3.2 Reinsurance layers

The model $L = \Omega^T\mathbf{X}$ can describe *gross* or *ground-up* losses in catastrophe risk assessment, but a bit more is needed to analyze reinsurance. Say a treaty covers the losses up to a limit λ after a retention (attachment point) of α and a co-reinsurance of κ. Specifically, define a reinsurance function as $r(L,\alpha,\lambda,\kappa) = (1-\kappa)\cdot \max(0, \min(L-\alpha,\lambda))$. This makes the net loss, after reinsurance, equal to $N = L - r(L,\alpha,\lambda,\kappa) = \Omega^T \cdot \mathbf{X} - r(\Omega^T \cdot \mathbf{X},\alpha,\lambda,\kappa) \equiv g(\mathbf{X};\theta)$. The symbol Ω is used to represent exposures here, because the parameter vector θ is the composite vector $\langle \Omega_1, \Omega_2, \ldots, \Omega_N, \alpha, \lambda, \kappa\rangle^T$.

Now the question of risk measures and gradients of risk measures of the net loss random variable N can be addressed.

3.2.1 Moments

The moment theorems still apply. While g is no longer smooth (its first derivatives are undefined at the bottom and top of the layer), it is piecewise smooth.

$$\nabla_\theta \mu[L] = E[\nabla_\theta g(\mathbf{X},\theta)] \text{ and } \nabla_\theta \sigma^2[L] = 2\cdot E[(g(\mathbf{X},\theta)-\mu)\cdot \nabla_\theta g(\mathbf{X},\theta)] \text{ and}$$

$$\nabla_\theta \sigma[L] = E\left[\frac{g(\mathbf{X},\theta)-\mu}{\sigma}\cdot \nabla_\theta g(\mathbf{X},\theta)\right].$$

Note, however, that the gradients of g are a bit different than in the bilinear case.

The gradient of net loss with respect to a change to exposure (for a given damage scenario **X**) is given by

$$\nabla_{\Omega}\, g = \begin{cases} \kappa \cdot \mathbf{X}, & \alpha < \mathbf{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \mathbf{X}, & otherwise \end{cases}.$$

That is, it is the damage rate at that location if the reinsurance is not "in the layer," but only κ times that if the contract is "in the layer."[7]

The partial derivative of net loss with respect to the treaty's attachment point is given by:

$$\frac{\partial g}{\partial \alpha} = \begin{cases} (1-\kappa), & \alpha < \mathbf{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ 0, & otherwise \end{cases}.$$

For every dollar that the layer is raised, the net loss increases by $(1-\kappa)$ dollars, but only if the contract is "in the layer."

The partial derivative of net loss with respect to the limit is given by:

$$\frac{\partial g}{\partial \lambda} = \begin{cases} (\kappa-1), & \alpha + \lambda < \mathbf{\Omega}^T \cdot \mathbf{X} \\ 0, & otherwise \end{cases}.$$

For every dollar the limit is increased, the net loss decreases by $(1-\kappa)$ dollars, but only if the contract has exhausted (paid out fully).

The partial derivative of net loss with respect to the co-reinsurance is given by:

$$\frac{\partial g}{\partial \kappa} = \begin{cases} 0, & \mathbf{\Omega}^T \cdot \mathbf{X} < \alpha \\ \mathbf{\Omega}^T \cdot \mathbf{X} - \alpha, & \alpha < \mathbf{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \lambda, & \alpha + \lambda < \mathbf{\Omega}^T \cdot \mathbf{X} \end{cases}.$$

Increasing co-reinsurance has no effect until the layer attaches, after which it applies to the amount eligible for reimbursement.

3.2.2 Exceedance probabilities

The application of the "integral over the surface formula" is not stymied by the discontinuous derivatives; one can divide the problem into three regions (below, in, and above the layer) where smoothness prevails. The gradients of Q with respect to N and the components of θ are computed according to the expectation formulas given in section 2.3.3. The gradient of g with respect to the damage variable **x** is given by

$$\nabla_{\mathbf{x}}\, g(x,\theta) = \begin{cases} \kappa \cdot \mathbf{\Omega}, & \alpha < \mathbf{\Omega}^T \cdot \mathbf{X} < \alpha + \lambda \\ \mathbf{\Omega}, & otherwise \end{cases}.$$

[7] Actually, the derivative is undefined in the cases where the gross loss equals α or $\alpha+\lambda$, but that only occurs on a set with probability zero.

Increasing damage rates are borne fully by the reinsured outside the layer, but only in proportion κ inside the layer.

Note, however, that if $\kappa = 0$, then the gradient $\nabla_{\mathbf{x}} g(\mathbf{x}, \boldsymbol{\theta})$ vanishes in the layer, making integrals involving $\|\nabla_{\mathbf{x}} g(\mathbf{x}, \boldsymbol{\theta})\|^{-1}$ nonexistent. In such a case, the probability distribution of N has a "mass point" at N=α, so the left-hand derivative of Q with respect to N, and the right-hand derivative of Q with respect to α, are both infinite there. The gradients of Q become undefined.

3.2.3 Value at Risk and Tail Value at Risk

With the details given in section 3.2.1 for gradients of g above, in, and below the layer, the computation of gradients of VaR and TVaR are straightforward applications of the formulas in 2.3.4-2.3.5. In the case where $\kappa = 0$, the exceedance probability Q is discontinuous and therefore the gradients of VaR and TVaR may be undefined.

3.3 Portfolio of layers

Here the canonical reinsurance pricing problem is addressed.[8] Consider a reinsurer who holds a portfolio of layers $L_c = r\left(\left(\boldsymbol{\Omega}^{<c>}\right)^T \cdot \mathbf{X}, \alpha_c, \lambda_c, \kappa_c\right)$. Here the symbol c indexes contracts, and the superscript <c> is used to denote vector #c (whereas a subscript would denote component #c of the vector).[9]

Ideally, pricing a proposed new layer L_n or considering the renewal of an existing layer L_e would take into account the effect of the change on the capital needs of the firm. The capital needs, in turn, would be driven (at least in part) by the risk characteristics of the portfolio $\{L_c\}$. A risk measure such as VaR or TVaR is typically used to make this assessment.

Reinsurers routinely compute the probability distribution of their payouts (and, hence, VaR and TVaR) by running a catastrophe model. This is usually done quarterly because the process is time-consuming.[10] How is one to assess the impact on VaR of adding a proposed new contract or nonrenewing an existing contract?

To cover either case, refer to the contract under consideration as contract #u. Examining contract u's contribution $\boldsymbol{\Omega}^{<u>}$ to the total exposure vector $\Sigma_c\, \boldsymbol{\Omega}^{<c>}$ is not helpful because of the nonlinear payoff of each contract – the total exposure $\Sigma\, \boldsymbol{\Omega}$ is virtually meaningless. Nonetheless, the gradient theorem for a risk measure R can be profitably applied to the following:

$$g(\mathbf{X};\theta) \;=\; \sum_c r\left(\left(\boldsymbol{\Omega}^{<c>}\right)^T \cdot \mathbf{X}, \alpha_c, \lambda_c, \kappa_c\right) \;+\; \theta \cdot r\left(\left(\boldsymbol{\Omega}^{<u>}\right)^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u\right)$$

where θ is a scalar equal to 1 for the case of new business u = n and -1 for the case of renewals u = e. The partial derivative with respect to θ will reveal (through the conditional expectation theorems) the impact on R of adding or dropping a "small amount" θ of the contract under consideration to or from the portfolio. If

[8] It also corresponds to the evaluation of a portfolio of call options at the (future) time of their exercise.

[9] Of course, real reinsurance often involves other complicating factors as well, such as reinstatements, per-occurrence and aggregate retentions and limits, complex "towers" of layers, backup capacity, etc., etc. The abstraction presented here is complex enough to make the point.

[10] Major and Kreps [2002] cite an example of seven hours.

the contract is already "small" compared to the portfolio, then the approximation
$\mathrm{R}\ |_{\theta=1} = \mathrm{R}\ |_{\theta=0} + \partial \mathrm{R}/\partial\theta$ will be "good."

The partial derivative of g with respect to θ

$$\frac{\partial}{\partial\theta} g(\mathbf{X};\theta) = \quad r\left(\left(\Omega^{<u>}\right)^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u\right)$$

is simply the payout of the contract under consideration. The gradient with respect to **X** is more complex, however.[11] The Monte Carlo section presented next will discuss an approach to evaluating $E_h[\nabla_\theta g]$ that does not require computing $\nabla_{\mathbf{x}} g$.

4. Monte Carlo implementation

4.1 MC simulation and computation of risk measures

The most general applications of risk measures involve complicated nonlinear loss functions g and relatively intractable probability distributions f, and are usually solved by Monte Carlo methods. The basic method is to generate a sample of risk factor vectors $\mathbf{X}^{<i>}$ (i = 1, 2, ..., 10000, say) and apply the loss function to them sequentially to create a corresponding sample of loss values $L_i = g(\mathbf{X}^{<i>};\boldsymbol{\theta})$. Statistical methods are used to estimate percentage points of the distribution of L, hence risk measures.

Variance reduction techniques (VRTs), usually involving special steps in the construction of the sample $\mathbf{X}^{<i>}$, can be applied to improve the accuracy/speed tradeoff. The most common VRTs create samples where the points are not equally probable and adjust the estimation procedure accordingly. Stratified sampling and importance sampling are two such approaches.[12] If each scenario $\mathbf{x}^{<i>}$ has an associated probability p_i, then one can approximate an expectation integral by a summation as follows:

$$\int \varphi(\mathbf{x}) \cdot f(\mathbf{x}) dV \approx \sum_i \varphi(\mathbf{x}^{<i>}) \cdot p_i$$

4.1.1 Moments

For example, the first and second moments of a loss $L = g(\mathbf{X};\boldsymbol{\theta})$ would be estimated by

$$\mu[L] = \int g(\mathbf{x};\boldsymbol{\theta}) \cdot f(\mathbf{x}) dV \approx \sum_i L_i \cdot p_i \quad \text{and}$$

$$\sigma^2[L] = \int (g(\mathbf{x};\boldsymbol{\theta}) - \mu)^2 \cdot f(\mathbf{x}) dV \approx \sum_i (L_i - \mu)^2 \cdot p_i$$

---

[11] Technically, there is a problem in that $\nabla_{\mathbf{x}} g$ vanishes where none of the layers attach or where all of the layers have exhausted. As long as the isoclast is in neither region, however, this does not present a practical problem.

[12] The generation of complementary pairs of sample points ("antithetic variates") while useful in estimating means, is generally not so helpful in tail-sensitive risk measure computations. See Rubinstein [1981] for more on the subject.

4.1.2 Exceedance probabilities

The exceedance probabilities are estimated by $Q(l;\theta) \approx \sum_{\{i:L_i>l\}} p_i$.

4.1.3 Value at Risk and Tail Value at Risk

The condition that defines the VaR level L_q, $q = Q(L_q;\theta)$, is evaluated by a search procedure. In practice, this means computing L_i for each scenario i, sorting the results, and then finding that position in the sort order where the sum of probabilities p_i above that point adds up to the desired exceedance probability q. The numerical value of that cut point is then the VaR L_q.

Having divided scenarios i into those above and those below the VaR, one can then estimate the TVaR as:

$$T_q \approx \frac{1}{q} \cdot \sum_{\{i:L_i>L_q\}} L_i \cdot p_i$$

4.2 MC simulation and computation of gradients

The gradient of a risk measure could be obtained by rerunning the model and recomputing the measure after making small changes to each portfolio parameter, in turn.[13] For problems of realistic size, involving hundreds or thousands of parameters (asset classes, insured locations, etc.) and taking hours to run a single risk measure computation, this approach would be impractical. A more efficient solution is to appeal to the appropriate expectation theorem.

4.2.1 Gradients of moments

The first and second moment gradients, because they are expressed as expectations, are easily computed via Monte Carlo:

$$\nabla_\theta \mu[L] = \int \nabla_\theta g(\mathbf{x};\theta) \cdot f(\mathbf{x}) dV \approx \sum_i \nabla_\theta g(\mathbf{x}^{<i>};\theta) \cdot p_i \quad \text{and}$$

$$\nabla_\theta \sigma^2[L] = 2 \cdot \int (g-\mu) \cdot \nabla_\theta g \cdot f(\mathbf{x}) dV \approx \sum_i \left(g(\mathbf{x}^{<i>};\theta) - \mu\right) \cdot \nabla_\theta g(\mathbf{x}^{<i>};\theta) \cdot p_i$$

The continuous (theoretical) versions follow from the fact that smoothness allows us to interchange differentiation and integration. For the discrete (Monte Carlo) version, no such justification is needed. Differentiation is linear, hence can be interchanged with summation and multiplication by constants.

4.2.2 Gradient of exceedance probabilities

The situation is not so bright with the probability gradient theorems. Strictly speaking, they do not apply to simulation models of this type. The probability distribution for **X** is discrete, so the isoclast of **X** values solving $g(\mathbf{X};\theta) = L_q$ is not a smooth manifold.[14] However, if the discrete distribution were modified by

---

[13] It would be helpful to use the same sample of **X** for each run, thereby reducing the between-run variance.

[14] Depending on how the quantile is defined, there may be only one or no points $x^{<i>}$ at all in the solution set.

kernel smoothing (Silverman [1986]), so that it became smooth, then the theorems could apply. In particular, each point in the discrete distribution of **X** should be replaced by a patch of density sufficiently large and smooth so that there are no "holes" in the resulting kernel mixture.

Using a normal kernel would accomplish this. Unfortunately, a normal kernel has infinite support, and would give every one of the original points $\mathbf{x}^{<i>}$ some nonzero contribution to the conditional distribution on the isoclast. Therefore, all points would need to be included in the computation of the gradient, even though most would be "far away" from the isoclast, contributing a negligible amount to the expectation. The computations would be needlessly burdensome.

For this reason, a kernel with finite support is recommended. By varying the kernel bandwidth, one can assess the numerical stability of the gradient computation. In particular, one can use the Epanechnikov kernel, defined as:

$$K(t,w)=\begin{cases}0, & |t|>w\\ 1-(t/w)^2, & |t|\le w\end{cases}$$

where t is the distance from the kernel origin and w is the kernel bandwidth.

The best way to understand why kernel smoothing makes sense in this context is to write the derivatives in the form of their definitions as limits of ratios. For example,

$$\frac{\partial Q}{\partial L}=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\int_{g(\mathbf{x})>L+\varepsilon} f(\mathbf{x})dV-\int_{g(\mathbf{x})>L} f(\mathbf{x})dV\right)=-\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{L<g(\mathbf{x})\le L+\varepsilon} f(\mathbf{x})dV .$$

The last expression, the N-dimensional volume integral on the hypershell $\{\mathbf{x} \mid L<g(\mathbf{x};\boldsymbol{\theta})<L+\varepsilon\}$ has "thickness" $\|\nabla_{\mathbf{x}} g\|^{-1}\cdot\varepsilon$ around the $g(\mathbf{x};\boldsymbol{\theta})=L$ hypersurface. This is why the limiting expression is $\partial Q/\partial L=-\int_{g(\mathbf{x},\theta)=L}\|\nabla_{\mathbf{x}} g(\mathbf{x},\boldsymbol{\theta})\|^{-1}\cdot f(\mathbf{x})dS$. If a finite ε were used in a Monte Carlo approximation of the hypershell integral, one would get

$$\frac{\partial Q}{\partial L}\approx-\frac{1}{\varepsilon}\sum_{i:L<g\le L+\varepsilon} p_i=-\frac{1}{\varepsilon}\sum_i I\left(\mathbf{x}^{<i>},\varepsilon\right)\cdot p_i$$

where the indicator function I takes on values 1 or 0 according to whether the point $x^{<i>}$ puts g inside the ε-hypershell or not. Taking the next step and using the Epanechnikov kernel, write instead

$$\frac{\partial Q}{\partial L}\approx-\frac{3}{4\cdot w}\cdot\sum_i K(L_i-L,w)\cdot p_i$$

where now w takes on the role that ε had. (The prefactor is the inverse of the expected value of K, 2/3, times the interval width 2w.)

Similarly, the partial derivative with respect to a component θ is approximated by

$$\frac{\partial Q}{\partial \theta}\approx\frac{3}{4\cdot w}\cdot\sum_i K(L_i-L,w)\cdot\frac{\partial g}{\partial \theta}\cdot p_i .$$

The numerical implementation here is, in effect, approximating the well-behaved hypershell limit integral by using an ε of the same order of magnitude as the kernel window. Note that the factor $\|\nabla_x g\|^{-1}$ does not explicitly occur in the simulation approach; rather, its effects are handled in the windowing.

In these summations, only points $x^{<i>}$ whose g-values L_i are within the kernel bandwidth of the VaR value need be included. This can yield a substantial efficiency improvement over methods that must evaluate all points.

4.2.3 Gradients of Value at Risk and Tail Value at Risk

The gradient of Value at Risk is computed directly as the ratio of the probability gradients derived previously.

$$\nabla_\theta \mathsf{L}_q = -\frac{\nabla_\theta Q}{\partial Q/\partial L} = \frac{\sum_{i:|L_i - \mathsf{L}_q|<w} K(L_i - \mathsf{L}_q, w)\cdot \nabla_\theta g \cdot p_i}{\sum_{i:|L_i - \mathsf{L}_q|<w} K(L_i - \mathsf{L}_q, w)\cdot p_i}.$$

The steps are as follows:

(1) Run the model once, saving an index to L_i values, and compute VaR as the desired empirical percentile.

(2) Determine which points i are within the kernel bandwidth.

(3) Run the model a second time, computing the vectors of partial derivatives $\partial g/\partial\theta$ at the $\mathbf{x}^{<i>}$ points identified in step 2.

(4) Tabulate the weighted average of the partial derivative vectors to obtain the gradient of VaR $\nabla_\theta \mathsf{L}_q$.

The gradient of TVaR is estimated as

$$\nabla_\theta \mathsf{T}_q(\theta) = E[\nabla_\theta g(\mathbf{X},\theta) \mid g(\mathbf{X},\theta) > \mathsf{L}_q(\theta)] = \frac{1}{q}\cdot \sum_{i:L_i > \mathsf{L}_q} \nabla_\theta g \cdot p_i$$

Note again the similarity between VaR and TVaR. In the Monte Carlo implementation, both are weighted averages of $\nabla_\theta g$; they differ in the x-sets over which the summation is carried out and in the weights themselves. VaR involves points **x** near the isoclast and with weights proportional to $K(w)p_i$. TVaR involves points **x** on one side of the isoclast and with weights proportional to p_i.

4.3 Computational savings over brute force approach

One can get a rough idea,[15] at least in relative terms, of how long it would take a computer to calculate VaR or TVaR. Assume that most of the work is in the computation of losses for one scenario $g(x_i,\theta)$ and this takes, on average, time T_g. Assume there are I scenarios (i=1,2,...,I), so the total time taken in

---

[15] The practical implementation issues of disk caching, overhead of function calls, etc., make this only a rough approximation.

computation of losses is $T_g \times I$. Sorting algorithms typically take time proportional to the log of the size of the list, say $T_s \times \log(I)$. A (relatively) fast linear accumulation search finds the VaR. Assume this time is negligible. The additional step in calculating TVaR consists of multiplying some number $q' \times I$ pairs of losses and probabilities.[16] This is about twice the work as the accumulation step for VaR, so this additional time will also be assumed negligible. To a first approximation, the time it takes to compute VaR or TVaR is $T_g \times I + T_s \times \log(I)$. If I is very large and the computation of g is particularly complex, this is approximately $T_g \times I$.

What about the gradients? Gradient components are given by

$$\frac{\partial}{\partial \theta_j} \mathrm{R}_q \approx \sum_{\{i\}} k_i \frac{\partial}{\partial \theta_j} g(\mathbf{x}^{<i>}, \boldsymbol{\theta}) \cdot p_i$$

where the weights k_i and the summation set $\{i\}$ differ according to whether R is VaR or TVaR. Say there are $q' \times I$ terms in the summation for TVaR and $q'' \times I$ terms for VaR. (Presumably, $q'' < q' << 1$.) Depending on the nature of the particular parameter θ_j, computing the partial derivative at a particular scenario may be difficult or easy. Worst case, it can be computed as

$$\frac{\partial}{\partial \theta_j} g(\mathbf{x}^{<i>}, \boldsymbol{\theta}) \approx \frac{g(\mathbf{x}^{<i>}, \boldsymbol{\theta} + \Delta\theta_j) - g(\mathbf{x}^{<i>}, \boldsymbol{\theta})}{\Delta\theta_j}$$

making it cost approximately $2 \times T_g$ in computer time. If N components j are to be evaluated, the total is $(N+1) \times T_g$ because the base case $g(\mathbf{x}^{<i>}, \boldsymbol{\theta})$ needs only be evaluated once and stored.

Say, in general, one evaluation cost $d \times T_g$. The better cases have very small d. For example, as was seen above, an exposure change is typically linear in the loss. So is a change to a pro-rata share. Limits and retentions on XOL contracts result in dollar-for-dollar changes or no change at all, depending on whether the scenario is below, in, or above the layer. In all of these examples, no further evaluation of g is needed – although there may be some overhead in tracking the status of the original g computation – and d is close to zero.

The "brute force" method of assessing changes in VaR (resp., TVaR) is to start with a base case and then re-evaluate it for each of N perturbations of the $\boldsymbol{\theta}$ parameter, for a total computation time of approximately $T_g \times I \times (1+N)$. In contrast, by using the gradient, this time can be cut to approximately $T_g \times I \times (1 + q^0 \times d \times N)$ where q^0 is q' (resp., q''). If N is large, then the ratio of time taken by using the theorems to the time taken by the brute force method is approximately $q^0 \times d$.

5. Analytical solutions and approximations

5.1 Delta-normal model

Perhaps the most widely-used analytical model for VaR, very familiar to financial risk management, is the Delta-normal model alluded to in section 3.1.4. The delta approximation writes:

$$L = g(\mathbf{x}_0 + \mathbf{X}) = g(\mathbf{x}_0) + (\nabla_x g) \cdot \mathbf{X}$$

[16] If scenarios are equally probable, then q'=q. If scenarios are oversampled at the high end, then q' is likely to be larger than q, but still less than one.

and considers the vector $\boldsymbol{\delta} = (\nabla_{\mathbf{x}} g)^T$ to comprise the portfolio parameter $\boldsymbol{\theta}$. When $\mathbf{X}$ is assumed to possess a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}$, then L has a (univariate) normal distribution with moments

$$E(L) = g(\mathbf{x}_0), \qquad \sigma^2(L) = \boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$$

Exceedance probabilities can be computed from the cumulative normal distribution Φ (Johnson et. al. [1994]) as:

$$Q(L) = 1 - \Phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right)$$

VaR is readily calculated from the appropriate percentile z_q of the normal distribution:

$$\mathsf{L}_q(\boldsymbol{\delta}) = E(L) + z_q \cdot \sqrt{\sigma^2(L)} = g(\mathbf{x}_0) + z_q \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}$$

To develop TVaR, a lemma about the conditional expectation of a normal random variable is needed.

Lemma: Given a scalar random variable Y distributed normally with mean μ and variance σ^2, then

$$E[\max(0, Y - t)] = \int_t^{\infty} (y - t) \cdot f(y) dy =$$

$$\xi(t, \mu, \sigma) \equiv \sigma \cdot \phi\left(\frac{t - \mu}{\sigma}\right) - (t - \mu) \cdot \left(1 - \Phi\left(\frac{t - \mu}{\sigma}\right)\right)$$

where ϕ is the standard normal probability density function and Φ is the standard normal cumulative density function. $\xi(t, \mu, \sigma)$ is related to the *limited expected value* function (Klugman et. al. [1998]).

Proof of Lemma. First, note that $\int_t^{\infty} z \cdot \phi(z) dz = \phi(t)$, which follows by differentiating $\phi(t)$.

Change variables from Y to $z = \dfrac{Y - \mu}{\sigma}$ and write the expectation as

$$\int_{\frac{t-\mu}{\sigma}}^{\infty} (\sigma \cdot z + \mu - t) \cdot \phi(z) dz = \sigma \cdot \int_{\frac{t-\mu}{\sigma}}^{\infty} z \cdot \phi(z) dz + (\mu - t) \cdot \int_{\frac{t-\mu}{\sigma}}^{\infty} \phi(z) dz$$

$$= \sigma \cdot \phi\left(\frac{t - \mu}{\sigma}\right) - (t - \mu) \cdot \left(1 - \Phi\left(\frac{t - \mu}{\sigma}\right)\right)$$

QED.

TVaR is then obtained as $\mathsf{T}_q = \frac{1}{q} \cdot \int_{\mathsf{L}_q}^{\infty} L \cdot f(L)dL = \mathsf{L}_q + \frac{1}{q} \cdot \xi\left(\mathsf{L}_q, g(\mathbf{x}_0), \sqrt{\delta^T \cdot \Sigma \cdot \delta}\right)$, which simplifies to $\mathsf{T}_q = g(\mathbf{x}_0) + \frac{\phi(z_q)}{q} \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}$.

5.1.1 Gradients in the delta-normal model

Above, the symbol $\boldsymbol{\delta}$ was used instead of $\boldsymbol{\theta}$ to emphasize that its nature is more in describing the portfolio than in controlling it. One cannot change the "delta" of a portfolio directly; it comes about as the result of some other sort of action (e.g., changing the amount of a risk held in the portfolio). Nonetheless, it is meaningful to inquire into the relationship between various risk measures and the delta of the portfolio.

The $\boldsymbol{\delta}$-gradients of the first two moments and the partial derivative of exceedance probability with respect to loss amount follow directly from the bilinear case examined in section 3.1:

$$\nabla_{\boldsymbol{\delta}} \mu = 0$$
$$\nabla_{\boldsymbol{\delta}} \sigma^2 = 2 \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$$
$$\frac{\partial Q}{\partial L} = -\frac{C_N}{\|\boldsymbol{\delta}\|}$$

Normality assumptions add nothing except to specify[17]

$$C_N = \int_{g(\mathbf{x})=L} f(\mathbf{x}) dS = \frac{\|\boldsymbol{\delta}\|}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}\right)$$

The gradient of exceedance probability with respect to $\boldsymbol{\delta}$ can be derived in two ways. Differentiation of the delta-normal formula for Q gives the gradient with respect to the $\boldsymbol{\delta}$ vector. An alternative is to apply the conditional expectation theorem. Both approaches are outlined in the appendix (section 9.3.1) and achieve the following result: $\nabla_{\delta} Q = \left(C_N / \|\boldsymbol{\delta}\|\right) \cdot E\left[\mathbf{X} \mid g(\mathbf{x}_0) + \boldsymbol{\delta}^T \cdot \mathbf{X} = L\right] = \frac{C_N}{\|\boldsymbol{\delta}\|} \cdot \frac{L - g(\mathbf{x}_0)}{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$.

Notice that the result is a vector because $\boldsymbol{\Sigma}$ is a matrix and $\boldsymbol{\delta}$ is a vector. The other terms in the product are scalars.

The gradient of VaR in the delta-normal model is then readily obtained as

$$\nabla_{\boldsymbol{\delta}} \mathsf{L}_q = \frac{z_q}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}.$$

Differentiating the expression for TVaR, or applying the gradient theorem, produces:

---

[17] To see this, rotate to a new orthonormal coordinate system $\mathbf{W}$ where $\mathbf{X} = W_1\mathbf{n}^{<1>} + W_2\mathbf{n}^{<2>} + \ldots + W_N\mathbf{n}^{<N>}$ and $\mathbf{n}^{<1>} = \boldsymbol{\delta}/\|\boldsymbol{\delta}\|$. Since $L = \|\boldsymbol{\delta}\| W_1$, it follows that the marginal pdf of W_1 is $\|\boldsymbol{\delta}\|$ times the marginal pdf of L.

$$\nabla_{\delta} \mathrm{T}_q = \frac{\phi(z_q)}{q} \frac{\Sigma \cdot \delta}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}$$

These are also detailed in the appendix (sections 9.3.2-9.3.3).

5.2 Gamma-normal model

It has been widely recognized that the delta-normal model does not do a good job when nonlinear instruments (e.g., call options) are present in the portfolio in significant quantities. That is, the first-order Taylor approximation to the loss function g is inadequate over the range of $\mathbf{X}$ values relevant to the computation of VaR or TVaR. To remedy this, various authors have suggested the use of the second-order Taylor expansion in the gamma-normal model. Following Britten-Jones and Schaeffer [1999], consider the model:

$$L = g(\mathbf{X}; \delta, \Gamma) = \nu + \delta^T \cdot \mathbf{X} + \frac{1}{2} \cdot \mathbf{X}^T \cdot \Gamma \cdot \mathbf{X} \qquad \mathbf{X} \sim N(\mu, \Sigma)$$

where δ is the vector of first-degree sensitivities and Γ is the (symmetric) matrix of second-degree sensitivities. $\mathbf{X}$ is assumed multivariate normal. This can be rewritten by "completing the square" to become:

$$L = \left(\nu - \frac{1}{2} \cdot \delta^T \cdot \Gamma^{-1} \cdot \delta \right) + \frac{1}{2} \cdot \sum_{i=1}^{N} \mathbf{D}_i \cdot (Z_i)^2 \qquad \mathbf{Z} \sim N\left(\mathbf{H}^T \cdot \left(\mu + \Gamma^{-1} \cdot \delta\right), \mathbf{I}\right)$$

where $\Sigma^{1/2} \cdot \Gamma \cdot \Sigma^{1/2} = \mathbf{H} \cdot \mathbf{D} \cdot \mathbf{H}^T$, $\mathbf{D}$ is a diagonal matrix of eigenvalues, $\mathbf{H}$ is orthonormal (its columns being eigenvectors), and $\mathbf{I}$ is the identity matrix. Thus L can be seen as a linear combination of non-central chi-square variables.

In particular, the moments can be expressed in closed form. The first two are:

$$E[L] = \nu + \delta^T \cdot \mu + \frac{1}{2} \cdot \mu^T \cdot \Gamma \cdot \mu + \frac{1}{2} \cdot trace(\Gamma \cdot \Sigma)$$

$$\sigma^2[L] = \delta^T \cdot \Sigma \cdot \delta + \frac{1}{2} \cdot (trace(\Gamma \cdot \Sigma))^2$$

Britten-Jones and Schaeffer [1999] discuss how to obtain a good, tractable, approximation to the cdf of L under some circumstances. Rouvinez [1997] numerically inverts the characteristic function of the distribution.[18]

5.3 Conditional (not necessarily normal) moment models

In computing the gradients of VaR or TVaR, it may be possible to obtain the conditional moments of X (μ_q, Σ_q) by simulation or analytical methods.

In particular, for VaR:

[18] Britten-Jones and Schaeffer also remark "It is always possible to proceed by simulation."

$$C = \sum_{i:|L_i - \mathrm{L}_q|<w} p_i \cdot K(L_i - \mathrm{L}_q, w)$$
$$\boldsymbol{\mu}_q = C^{-1} \cdot \sum_i \mathbf{x}^{<i>} \cdot p_i \cdot K(L_i - \mathrm{L}_q, w)$$
$$\boldsymbol{\Sigma}_q = C^{-1} \cdot \sum_i (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q) \cdot (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q)^T \cdot p_i \cdot K(L_i - \mathrm{L}_q, w)$$

and for TVaR:

$$C = \sum_{i:L_i > \mathrm{L}_q} p_i$$
$$\boldsymbol{\mu}_q = C^{-1} \cdot \sum_i \mathbf{x}^{<i>} \cdot p_i$$
$$\boldsymbol{\Sigma}_q = C^{-1} \cdot \sum_i (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q) \cdot (\mathbf{x}^{<i>} - \boldsymbol{\mu}_q)^T \cdot p_i$$

For a bilinear model (e.g., the delta model of sections 3.1 and 5.1) $\boldsymbol{\mu}_q$ can be substituted for the conditional mean of **X** in the VaR and TVaR gradient formulas (section 3.1.3); covariance is of no consequence.

A quadratic g function (the gamma model, specialized in section 5.2) is nearly as simple, because $g(\mathbf{X}) = \nu + \boldsymbol{\delta}^T \cdot \mathbf{X} + \frac{1}{2} \cdot \mathbf{X}^T \cdot \boldsymbol{\Gamma} \cdot \mathbf{X}$ implies $E[\nabla_{\boldsymbol{\delta}} g] = \boldsymbol{\mu}_q$ and $E[\nabla_{\boldsymbol{\Gamma}} g] = (1/2)\boldsymbol{\Sigma}$. This follows from term-by-term differentiation. If $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ are themselves functions of a scalar parameter θ, then $E\left[\frac{\partial g}{\partial \theta}\right] = \frac{\partial \boldsymbol{\delta}}{\partial \theta} \cdot \boldsymbol{\mu}_q + \frac{1}{2} \cdot \sum_{i,j} \left(\frac{\partial \Gamma_{i,j}}{\partial \theta} \cdot \Sigma_{i,j} \right)$. This follows from the chain rule. If $\boldsymbol{\theta}$ is a vector parameter, the obvious componentwise extension applies. Here, only the first two moments matter and normality is not needed.

5.4 Conditional normality yields analytic solution for XOL portfolio management

The problem of evaluating the marginal impact of adding or dropping a cat layer yields to a tractable solution under the assumption of conditional normality. Recall the marginal impact model from section 3.3:

$$g(\mathbf{X};\theta) \;=\; g_0 + \theta \cdot r\left((\boldsymbol{\Omega}^{<u>})^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u \right)$$

with the solution $\nabla_\theta \mathrm{R}_q = E_*\left[r\left((\boldsymbol{\Omega}^{<u>})^T \cdot \mathbf{X}, \alpha_u, \lambda_u, \kappa_u \right) \right]$ where the conditional expectation is taken with respect to the distribution appropriate for R being VaR or TVaR. If **X** is assumed conditionally distributed as $N(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$, define the following:[19]

---

[19] Computation of $\boldsymbol{\mu}_q$ and $\boldsymbol{\Sigma}_q$ are discussed in section 5.3.

$$\mu = \left(\Omega^{<u>}\right)^T \cdot \mu_q$$
$$\sigma^2 = \left(\Omega^{<u>}\right)^T \cdot \Sigma_q \cdot \Omega^{<u>}$$
$$\varsigma(t) = \xi(t, \mu, \sigma)$$

where $\xi(t,\mu,\sigma)$ was defined in section 5.1. Then:

$$\nabla_\theta \mathrm{R}_q = E_*[r_u] = (1-\kappa_u)\cdot(\varsigma(\alpha_u) - \varsigma(\alpha_u + \lambda_u)).$$

This follows from the lemma proved in section 5.1 and the construction of the r function.

The normal approximation reduces the calculation of the marginal impact of a contract to a matrix multiplication, two vector multiplications, and a relatively simple scalar formula.[20]

This technique can be refined by extending to a mixture of normals.[21] Divide sample points into groups $\{\mathbf{x}^{<i1>}\}, \{\mathbf{x}^{<i2>}\},\ldots, \{\mathbf{x}^{<ik>}\},\ldots, \{\mathbf{x}^{<iG>}\}$, e.g., by a clustering algorithm (Anderberg [1973]). Compute separate moment vectors μ_k and variance-covariance matrices Σ_k for each group k as described previously. Instead of a single multivariate normal with density $\phi(\mathbf{X};\mu,\Sigma)$, the mixture has density $\sum_{k=1}^{G} \omega_k \cdot \phi(\mathbf{X};\mu_k,\Sigma_k)$, where the coefficients ω_k are proportional to $\sum_{i_k} p_{i_k} \cdot K(L_{i_k} - \mathrm{L}_q, w)$ for the VaR and to $\sum_{i_k} p_{i_k}$ for the TVaR (and sum to one). The layer expectation is similarly a weighted sum of terms of the previous form. The added cost is more computation;[22] the benefit is improved accuracy of the approximation.

6. Case study

6.1 Background: CAT model

In this section a simplified CAT model illustrates the application of the gradient theorems and helps assess their numerical accuracy in the context of Monte Carlo simulation.

Consider a hypothetical stretch of coastline with locations measured from 0 to 10. An event will consist of a simulated hurricane making landfall at location t and producing damage at location s of $X(s,t,r)$ where the parameter r is the maximum loss rate for that event. Specifically,

---

[20] Good approximations for the normal cdf are available in many mathematics and statistics handbooks (e.g., Abramowitz and Stegun [1964], Johnson et. al. [1994]) and are built-in functions in many PC software packages, e.g. Excel.

[21] One could go further and model arbitrary g by a piecewise linear approximation built up from several r functions.

[22] Not necessarily proportional to G, however. It may be the case that separate groups reflect a number of shared zero X-components, thus reducing the effective dimensions of the group variance-covariance matrices compared to the matrix for the entire isoclast which must be at least as large.

$$X(s,t,r)=\frac{r}{1+(t-s)^2}.$$

Figure 1 shows such a damage profile for t = 0.4, r = 1.

The simulation will create random hurricanes (**X**-vectors indexed by *s*) according to a specific joint distribution for *t* and *r*. The joint distribution is defined as follows. The variable *t* is distributed uniformly between 0 and 10. Conditional on *t*, the variable *r* is distributed exponentially with mean

$$\bar{r}|t=\left(10+4\cdot\cos(0.2\cdot\pi\cdot(t-1.5))+2\cdot\sin(0.7\cdot\pi\cdot(t-4.5))\right)^{-1}$$

Figure 2 graphs the conditional mean of *r* as a function of *t*.

From this distribution, 50,000 equal-probability events were generated. This set was resampled by importance sampling (Rubinstein [1981]) on *r* resulting in 5002 events with associated sampling weights (relative frequencies). Figure 3 shows a 20% subsample of these events.

Events generated as *(r,t)* pairs were converted to vectors $<X_s: s = 0,1,\ldots,10>$ and stored on disk.

6.2 Ground up losses

Figure 4 shows, and Table 1 enumerates, a hypothetical schedule of insured property exposures along a stretch of coastline, with locations numbered from 0 to 10. This vector corresponds to the parameter $\mathbf{\Omega}$ in the equation $L=\mathbf{\Omega}^T\cdot\mathbf{X}$. Note the subscripting needs to handle potential as well as actual members of the portfolio. In a real simulation, rather than list thousands of individual insured properties one would create group aggregations, say, by ZIP code and occupancy type. If location #s has no current presence, but a potential future presence, in the portfolio being studied, then $\Omega_s = 0$, but there is still a need to model the risk factor component $\mathbf{X}_s$.

After carrying out the multiplication $L_i=\mathbf{\Omega}^T\cdot\mathbf{X}^{<i>}$ for each event *i*, one has a discrete distribution of total (ground up) losses L_i each with associated probability p_i. The cumulative distribution function is shown in figures 5a and 5b. Moments for the loss are $\mu = 0.682$ and $\sigma = 0.805$. The mean is indicated in figure 5a.

Take $q = 0.02$ as the reference point for VaR and TVaR. VaR for losses is 3.139, or just a bit more than 3 standard deviations above the mean. VaR is indicated in figures 5a and 5b. TVaR is 4.251.

The gradients of VaR and TVaR were computed both via the "brute force" approach and by the conditional expectation theorems (section 4.2.3). The brute force approach was to make a discrete change $\Delta\Omega_s$ to the exposure at location *s*, recomputing the statistic R, and then calculating the rate of change by $(R(\Omega+\Delta\Omega)-R(\Omega-\Delta\Omega))/2\Delta\Omega$. This requires a choice for the increment $\Delta\Omega$. Values of 2 and 8 were chosen for illustrative purposes.

The computation of ∇TVaR by the expectation theorem[23] does not require a "bandwidth." The results are shown in figure 6. The vertical axis (for the curves) represents the gradient of TVaR with respect to changes in the exposure at each location. Exposures are shown by small bars at the bottom of the graph. The results from the theorem are shown by the dashed line. The thick and thin lines show the brute force approach with different $\Delta\Omega$. The results are quite consistent between the three methods. Locations 6 and

---

[23] For ground-up loss, $\nabla_\Omega g=X$.

7, with the most intense event risk, are also the most sensitive to exposure change. Locations 3, 4, and 10 are the least. Location 1 is more sensitive than locations 9 and 10, despite the fact that the event risk at location 1 is less intense than it is at the right-hand end. This is because of the relative concentration of exposures on the left end. Of the 5002 events, 781 (15.6%) entered into the theorem-based calculation of ∇TVaR.

The computation of ∇VaR by the expectation theorem does require a "bandwidth" w. Choices of 0.05 and 0.2 were used.[24] The results are shown in figure 7. They are quite similar to that of the TVaR graph, except that the various methods and parameter picks show less consistency. For $w = 0.05$, only 66 events were within the kernel, and for $w = 0.2$, 213 events. Thus by using the expectation theorems only 1.3% or 4.3% of the events needed to enter into the weighted average. Since this is the bilinear model, the vectors being averaged took on a particularly simple form (i.e., $\nabla_\Omega g = \mathbf{X}$).

Coupled with a model for the gradient of profitability, the gradient of VaR or TVaR can map proposed (small) portfolio changes into changes in the risk-reward plane of a portfolio risk management model. For example, if the relative profitability of business varied little over the locations, strategies involving writing business in locations 2-4 and 10, possibly coupled with attrition at locations 1 and 6, would yield the best changes in risk-reward position. While some of this might seem obvious from examination of the exposures and damage distributions alone, in a real application there would be many more locations (and classes of business) with correspondingly more difficult tradeoffs to assess.

6.3 Reinsurance buying

To make the problem a bit more realistic, consider that each location has a per-risk treaty associated with it, with an attachment $\alpha_s = 0.5$, limit $\lambda_s = 1$, and co-reinsurance $\kappa_s = 0.1$.

The mean and standard deviation of net loss are 0.569 and 0.563, respectively. Thus the expected net loss is 84% of the expected gross loss. The 2% VaR is 2.00, only 64% of the gross loss's, and the TVaR is 2.98, about 70% of what it is for the gross loss.

But is this the optimal set of attachments and limits? After all, the exposures vary by a factor of 4:1 from location to location. Perhaps it would make more sense to alter the α and λ parameters. Specifically, how do changes in α and λ vectors translate to changes in VaR and TVaR? Here, the gradient theorems must be applied in their more general form, because the relationship between net loss and the parameters in question is not linear.[25] There are 12 nontrivial parameters to be assessed here: one attachment point and one limit for each of the six locations with a treaty. Attachments and limits for locations at which there is no exposure have no effect on the loss and hence a zero gradient.

Figure 8 shows the ∇TVaR results for the attachment parameters. Values for $\Delta\alpha$ of 0.05 and 0.2 were chosen for the brute-force method. Curiously, location 1 is much less sensitive to a change in attachment than the adjacent locations 0 and 2. Out of the tail events defining TVaR, location 1 is only "in the layer" about 4.4% of the time whereas locations 0 and 2 are in the layer 18% and 21% of the time, respectively. While location 1 is more likely to exhaust the layer, that situation is insensitive to changes in attachment point. In particular, this means that the attachment point can be raised with (relatively) little impact on the net loss. A canny buyer would investigate the possibility of a premium reduction to accompany such a change. For net losses, 812 events (16.2%) entered into the computation of ∇TVaR by the expectation method.

---

[24] These choices are consistent with the $\Delta\Omega$ picks of 2 and 8 because the mean X across all locations is 0.027.

[25] $\nabla_\alpha g$ and $\nabla_\lambda g$ are given in section 3.2.1.

Location 6 has a low sensitivity for the same reason as location 1: in the tail, location 6 usually either does not attach (42.9% of the time) or completely exhausts the layer (55.2%); it is only in the layer 1.7% of the time.

Figure 9 shows the comparable results for ∇VaR with respect to attachment parameters. Again, bandwidth w = 0.05 and 0.2 were chosen for the expectation method.[26] These results are comparable to those of the TVaR graph, and again show less consistency. For the two bandwidth picks, 78 (1.6%) and 338 (6.8%) events, respectively, entered into the expectation calculations.

Figures 10 and 11 show the results for ∇TVaR and ∇VaR with respect to changes in the limits at each location. Since increasing a limit can only reduce net loss, the negative of the gradients is graphed. Changing limit is most effective at location 6 (which exhausts its cover in over half the tail events) and nearly ineffective at location 2 (which only exhausts in about 1% of the tail events).

6.4 Reinsurer portfolio of XOL layers

Consider now the problem of assessing the impact that adding a new contract or nonrenewing an existing contract has on the VaR and TVaR of a portfolio of layers.

Table 2 shows a proposed contract and 25 existing contracts. Each is defined by exposures at the same 11 locations as modeled previously, with attachments and limits applying to the total damages from all locations of a contract combined. With this small an example, it is feasible to rerun the model on the new portfolio, to find the new values of VaR and TVaR, and hence the new contract's marginal contributions to risk. On the other hand, application of the theorems requires evaluating the losses on *only* the new contract, and *only* for the scenarios within the kernel window or tail. In this case, that means one instead of 25 contracts and typically 2% to 20% of the scenarios (depending on bandwidth and risk measure), resulting in hundreds of times fewer loss evaluations.

The mean payout for this portfolio is 0.68, and its standard deviation is 2.908. The mean payout of the proposed new contract is 0.033 and its standard deviation is 0.175. Figure 12 shows the means and standard deviations for the existing contracts, with the statistics for the new contract overlaid as horizontal dotted lines.

The 2% VaR for the existing portfolio is 12.48, a bit more than 4 standard deviations above the mean. The TVaR is 18.21. Of the 5002 simulated events, 825 (16.5%) are in the 2% upper tail.

Direct computation of the change in VaR or TVaR upon adding or dropping a contract does not require a choice for the change amount as in previous examples; the natural unit is one contract. The computation of ∇TVaR by the expectation theorem does not require a bandwidth.

The approximation formula for ∇TVaR (based on the model for g in section 3.3) is:

$$\nabla \mathrm{T}_q(\boldsymbol{\theta}) \approx \frac{1}{q} \cdot \sum_{\{i: L_i > \mathrm{L}_q(\boldsymbol{\theta})\}} r\left(\left(\boldsymbol{\Omega}^{<u>}\right)^T \cdot \mathbf{X}^{<i>}, \alpha_u, \lambda_u, \kappa_u\right) \cdot p_i$$

where the summation, please note, is over the scenarios where the loss to the *existing* portfolio exceeds the VaR. In the case of dropping existing contracts, this formula is exactly the discrete implementation of the *co-TVaR* measure (Kreps [2003]). In the case of adding a proposed contract, the slight difference here is that the VaR does not include the effect of the proposed contract.

---

[26] These are consistent with Δα because in the layers, changes in attachment are approximately one-for-one effective in changing net losses.

The change in TVaR upon adding the new contract is 0.682, whereas it is estimated by the expectation theorem to be 0.683. Figure 13 shows the change in TVaR upon cancelling each of the existing contracts. The estimates are quite close, with a root mean square error of 0.012.

For the computation of ∇VaR by the expectation theorem, bandwidths of 0.5 and 1.5 were assumed. These encompassed 74 and 245 events, respectively. The change in VaR by adding the new contract is 0.198; it is estimated as 0.246 or 0.234 for the two bandwidths, respectively, 24% and 18% higher than the actual. Figure 14 shows the change in VaR upon cancelling each of the existing contracts. The estimates track fairly well, with a root mean square error of 0.083.

6.5 Normal approximation

Using the mean and variance-covariance matrix of the **X** vectors in the 2% tail, one can assume a normal distribution and evaluate ∇TVaR by a closed-form expression (section 5.4). The estimate for the new contract is 0.728, about 7% higher than the actual figure of 0.682. Figure 15 shows the results for the existing contracts. The root mean square error is 0.092.

Again, less accurate results are obtained for closed-form ∇VaR computation. The estimates for the new contract based on the two bandwidths are 0.304 and 0.329, respectively, 53% and 66% higher than actual. Figure 16 shows the results for the existing contracts. These had root mean square errors of 0.121 and 0.114, respectively.

7. Related work

The Value-at-Risk (VaR) approach to risk assessment is becoming more prevalent in the insurance industry as computing power increases and the convergence between finance and insurance disciplines continues. However, the VaR concept is not new to insurance. Actuaries have been familiar with the VaR approach (although not the name) since at least 1969, when McGuinness [1969] proposed a mathematical definition for PML.[27] What is different today is the recognition of the similarities between insurance and finance risk management needs and the consequent cross-fertilization of methodologies for evaluating the risk of a portfolio (whether that be an investment portfolio, a portfolio of insurance policies, or an entire enterprise).

Kreps [1990] links reinsurance pricing to the incremental risk a new contract adds to the reinsurer's portfolio. Kreps [1999] places reinsurance pricing on the same theoretical footing as other investment decisions. Major and Kreps [2002] present a field study on how reinsurers think and behave with respect to pricing catastrophic risk.

Kreps [2003] provides a general method to decompose risk measures into additive component "co-measures." Tasche [1999] develops a theory of capital allocation and portfolio component returns where the first derivative of a (fairly arbitrary) risk capital measure of choice plays a central role. His principle of "suitability" can be interpreted as saying that the rate of return on a component of a portfolio (with respect to allocated capital) should exceed the portfolio rate of return in precisely those situations where an additional increment of that component would increase the overall portfolio rate of return. He shows that in order for suitability to hold, the capital allocation must be equal to the first derivative of the risk measure. Acerbi [2002] presents a "spectral decomposition" of coherent risk measures which includes TVaR as a special case. Major [2003] provides an analysis of suitability for a variety of risk measures and explores the relationship between suitability, spectral measures, and co-measures.

---

[27] Since VaR is essentially a quantile of the cumulative distribution function, it is difficult to assign original credit. Gauss's [1809] work on the normal distribution extended the work of De Moivre [1738] and Laplace [1783], with other distributions having been studied a century earlier.

The literature on VaR is enormous; many papers are accessible through www.gloriamundi.org. Jorion [2000] is a good introduction. Fuglsbjerg [2000] applies variance reduction techniques to estimate VaR. Glasserman et. al. [2000] discuss efficient VaR estimation in the presence of heavy-tail risk factors. Artzner et. al. [1999] present axioms for risk measures which lead many to conclude that TVaR is superior to VaR.

The gradient of Value at Risk in the context of the linear loss model is well known. Garman [1996] analyzes the gradient in the delta-normal model and provides a blueprint for using the gradient in risk management. Hallerbach [1999] focuses on the decomposition of VaR between the risk factors, with a nuanced discussion distinguishing marginal, component, and incremental VaR. Gourieroux et. al. [2000] also deal with the second derivative and discuss kernel smoothing in the context of estimation. Tasche [1999] presents the gradient in a more general case for the distribution of the risk factors. Wang [2002] develops the Monte Carlo sampling distribution of incremental VaR.

Many researchers have gone beyond the derivative to explore VaR and TVaR optimization in the linear loss model. Andersson et. al. [2001] apply their results to the problem of credit risk TVaR. Rockafellar and Uryasev [2000], Uryasev [2000a, b], Krokhmal et. al. [2001], and Larsen et. al. [2002] develop a technique that simultaneously calculates VaR and optimizes TVaR. Krokhmal et. al. [2001] provide an example based on a portfolio of S&P 100 stocks. Larsen et. al. [2002] provide an example based on a portfolio of emerging market bonds. Gaivoronski and Pflug [2000] optimize the VaR and show that mean-VaR-optimal frontiers are quite different from mean-TVaR frontiers. Lemus Rodriguez [1999] proposes a nonparametric statistical technique for the estimation of the gradient and uses it to numerically optimize linear combinations of VaRs. Yamai and Yoshiba [2002] focus on sampling issues in comparing VaR and TVaR optimization.

The literature on the gradient of VaR in the context of nonlinear loss models is less well developed. Britten-Jones and Schaefer [1999] "complete the square" in the quadratic model to show that the distribution of losses is a sum of non-central chi-squares. They also provide formulas for the mean and variance and show how to approximate the cumulative distribution function.[28] Uryasev [1995a, b; 1999] provides the general framework for evaluating derivatives as integrals for arbitrary loss models. Uryasev [2000a] explicitly treats the derivative of VaR in the general case. (The optimization discussed there is with regard to the linear model, however.) Major et. al. [2001] and Belubekian [2003] discuss the use of the gradient in an insurance portfolio management context.

There is also a vast body of literature on other approaches to risk measurement, analysis, and management. Only a few items will be mentioned here. Berger et. al. [1998] optimize risk-adjusted return through the "tabu search" technique. Manganelli et. al. [2002] develop the gradient of GARCH variance in the linear model. Studer [1995] and Studer and Luthi [1996] present "Maximum Loss" as an alternative concept to VaR, and show its relationship with and advantages to VaR.

8. Conclusion

The gradient of a risk measure is a powerful tool for analyzing the high-dimensional problem of directing change in the composition of a portfolio. Whereas thousands of simulation runs might be needed to test all the relevant "what-if" scenarios (assuming they could all be identified), the gradient can quickly point out the maximally advantageous directions for change, whether that means growth in opportunity areas, attrition in risk-concentrated areas, or adjusting other contractual terms. Since Parkinson's Law (Parkinson

[28] See section 5.2 for a discussion

[1958]) seems to apply to simulation modeling,[29] the ability to lever modeling resources in this manner is crucial to a thorough analysis.

This paper reviewed basic risk measures including VaR and TVaR and discussed various closed-form and simulation implementations of the gradient. Insurance examples illustrated their use in CAT modeling, reinsurance program design, and reinsurance portfolio management. The general simulation methods are equally effective with linear parameters such as exposure amounts and non-linear parameters such as retentions and limits. A normal approximation was shown to solve the XOL contract marginal impact problem with a simple formula.

9. Appendix

9.1 Dictionary of mathematical symbolism

9.1.1 Styles and fonts

These standards are typical, but not rigidly enforced:

- **bold** is a vector or matrix
- *italic* is a scalar variable or function
- CAPITAL is a random variable
- γρεεκ is a parameter
- `Special` is a probabilistic operator (risk measure)

9.1.2 Specific symbols

$\mathbf{X}$	random variable in R^N, represents hazard process (section 2.1)
$f(\mathbf{x})$	probability density function of **X** (section 2.1)
$L = g(\mathbf{X},\boldsymbol{\theta})$	defines the losses to a portfolio of insured exposures (section 2.1)
$\boldsymbol{\theta}$	parameter in R^M, describes the transformation of **X** to a loss dollar amount (section 2.1)
$\int_t f(\mathbf{x})dV$	volume integral where **x** ranges over the N-dimensional region *t* in R^N
$\int_t f(\mathbf{x})dS$	surface integral where **x** ranges over the (N-1)-dimensional submanifold *t* in R^N
$E[g(\mathbf{X})]$	mathematical expectation of g(**X**) with respect to pdf $f(\mathbf{x})$
$h(\mathbf{x})$	transformed density on isoclastic hypersurface (section 2.3.4)
$k(\mathbf{x})$	conditional density in tail (section 2.3.5)
$E_h[g(\mathbf{X})]$	mathematical expectation of g(**X**) with respect to pdf $h(\mathbf{x})$
$\mathsf{R}(\boldsymbol{\theta})$	generic risk measure (section 2.2.4)
L_q	value at risk or VaR (section 2.2.3)
T_q	tail value at risk or TVaR (section 2.2.3)

[29] "Work expands to fill the available time" becomes "requirements expand to absorb available computer resources."

$\partial f / \partial x$	partial derivative of f with respect to x (section 2.2.4)
∇	gradient operator; vector of partial derivatives (section 2.2.4)
$\lVert \mathbf{v} \rVert$	norm (magnitude) of vector **v**
C_N	normalizing factor, probability integrated over the isoclastic surface (section 2.3.3)
α	attachment (retention) of reinsurance layer (section 3.2)
λ	limit of reinsurance layer (section 3.2)
κ	co-reinsurance of reinsurance layer (section 3.2)
$\boldsymbol{\Omega}$	parameter vector defines exposures (section 3.2)
$r(L, \alpha, \lambda, \kappa)$	reinsurance payoff function (section 3.2)
$N = L - r$	net loss after reinsurance (section 3.2)
$\boldsymbol{\Omega}^{<c>}$	vector number c; column c of matrix (section 3.3)
$K(t,w)$	Epanechnikov kernel function (section 4.2.2)
$\boldsymbol{\mu}$	mean operator or mean of **X**
$\Phi(z)$	cumulative normal distribution function (section 5.1)
$\phi(z)$	normal probability distribution function (section 5.1)
$\boldsymbol{\mu}_q$	conditional mean of **X** at the VaR isoclast or TVaR tail (section 5.3)
$\boldsymbol{\Sigma}$	summation operator or variance-covariance matrix of **X**
$\boldsymbol{\Sigma}_q$	conditional variance-covariance of **X** at the VaR isoclast or TVaR tail (section 5.3)
$\xi(t, \mu, \sigma)$	expected value E[max(0,Y-t)] for normal distribution (section 5.1)

9.2 Gradient of Tail Value at Risk

The gradient of TVaR is derived. Recall

$$\mathrm{T}_q(\boldsymbol{\theta}) = \frac{1}{q} \cdot \int_{g(\mathbf{x},\boldsymbol{\theta}) > \mathrm{L}_q(\boldsymbol{\theta})} g(\mathbf{x}, \boldsymbol{\theta}) \cdot f(\mathbf{x}) dV .$$

Applying the integral over the surface formula with $\boldsymbol{\psi} = \boldsymbol{\theta}$, $\mathrm{H} = \mathrm{T}_q$, $\phi = g(\mathbf{x},\boldsymbol{\theta})\mathrm{f}(\mathbf{x})$, and $\gamma = \mathrm{L}_q - g(\mathbf{x},\boldsymbol{\theta})$,

$$\nabla_\theta \mathrm{T}_q(\theta) = E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathrm{L}_q\right] + \frac{1}{q} \cdot \int_{g(\mathbf{x},\theta)=\mathrm{L}_q} \frac{\nabla_\theta\left(g(\mathbf{x},\theta) - \mathrm{L}_q\right)}{\left\|\nabla_\mathbf{x} g(\mathbf{x},\theta)\right\|} \cdot g(\mathbf{x},\theta) \cdot f(\mathbf{x}) dS =$$

$$= E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathrm{L}_q\right] + \mathrm{L}_q \cdot \frac{C_N}{q} \cdot E\left[\frac{\nabla_\theta\left(g(\mathbf{X},\theta) - \mathrm{L}_q\right)}{\left\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\right\|} \middle| g(\mathbf{X},\theta) = \mathrm{L}_q\right]$$

$$where \quad C_N = \int_{g(\mathbf{x},\theta)=\mathrm{L}_q} f(\mathbf{x}) dS.$$

$$= E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathrm{L}_q\right] + \mathrm{L}_q \cdot \frac{C_N}{q} \cdot \left(E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\left\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\right\|} \middle| g = \mathrm{L}_q\right] - \nabla_\theta \mathrm{L}_q \cdot E\left[\frac{1}{\left\|\nabla_\mathbf{x} g(\mathbf{X},\theta)\right\|} \middle| g = \mathrm{L}_q\right]\right)$$

Recall (section 2.3.4)

$$\nabla_\theta \mathrm{L}_q = \frac{E\left[\frac{\nabla_\theta g(\mathbf{X},\theta)}{\left\|\nabla_\mathbf{X} g(\mathbf{X},\theta)\right\|} \mid g(\mathbf{X},\theta) = \mathrm{L}_q\right]}{E\left[\frac{1}{\left\|\nabla_\mathbf{X} g(\mathbf{X},\theta)\right\|} \mid g(\mathbf{X},\theta) = \mathrm{L}_q\right]}$$

Therefore the expression in parentheses is zero and we are left with

$$\nabla_\theta \mathrm{T}_q(\theta) = E\left[\nabla_\theta g(\mathbf{x},\theta) \middle| g(\mathbf{X},\theta) > \mathrm{L}_q\right].$$

9.3 Derivations of gradients under the delta-normal model

9.3.1 Gradients of exceedance probability

Here, the gradients $\frac{\partial Q}{\partial L} = -\frac{C_N}{\|\delta\|}$ and $\nabla_\delta Q = \frac{C_N}{\|\delta\|} \cdot \frac{L - g(\mathbf{x}_0)}{\delta^T \cdot \Sigma \cdot \delta} \cdot \Sigma \cdot \delta$, presented in section 5.1.1, are derived in two ways.

Method 1: Differentiation of the expression $Q(L;\delta) = 1 - \Phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right)$

$$\frac{\partial Q}{\partial L} = -\phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right) \cdot \frac{\partial}{\partial L}\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right) = -\frac{1}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right)$$

$$\nabla_\delta Q = -\phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right) \nabla_\delta \left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right) = \frac{L - g(\mathbf{x}_0)}{\left(\delta^T \cdot \Sigma \cdot \delta\right)^{3/2}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right) \cdot \Sigma \cdot \delta$$

Recalling (section 5.1.1) that $C_N = \frac{\|\delta\|}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \phi\left(\frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}}\right)$, the expressions using C_N follow.

Method 2: Using the conditional expectation theorem for Q (section 2.3.3), provides insight into the nature of the isoclast. As a consequence of the linear case,

$$\frac{\partial Q}{\partial L} = -C_N \cdot E\left[\|\delta\|^{-1} \middle| \delta^T \cdot \mathbf{X} = L - g(\mathbf{x}_0)\right] = -C_N / \|\delta\|.$$

$$\nabla_\delta Q = C_N \cdot E\left[\frac{\nabla_\delta g}{\|\nabla_x g\|} \middle| g = L\right] = C_N \cdot E\left[\frac{\mathbf{X}}{\|\delta\|} \middle| \delta^T \cdot \mathbf{X} = L - g(\mathbf{x}_0)\right].$$

Apply a singular value decomposition $\Sigma = \Lambda \cdot \mathbf{P} \cdot \Lambda^T$ where $\mathbf{P}$ is diagonal (eigenvalues) and Λ (eigenvectors) satisfies $\Lambda^T\Lambda = \mathbf{I}$. Then transform variables to $\mathbf{X} = \Lambda \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z}$ where $\mathbf{Z}$ is a spherical unit normal (i.e. mean zero and identity variance-covariance matrix) random variable. Therefore, one can rewrite

$$\nabla_\delta Q = \frac{C_N}{\|\delta\|} \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot E\left[\mathbf{Z} \mid \delta^T \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z} = L - g(\mathbf{x}_0)\right].$$

The conditional distribution of $\mathbf{Z}$ is an N-1 dimensional spherical normal with mean=mode being the point $\mathbf{z}_0$ on the conditional hyperplane closest to the origin. The unit vector $\mathbf{u}$ in that direction is given by:

$$\mathbf{u} = \frac{\mathbf{P}^{1/2} \cdot \Lambda^T \cdot \delta}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}}$$

Putting $\mathbf{z}_0 = k\mathbf{u}$ and solving for k gives us $k = \frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}}$ and therefore

$$\nabla_\delta Q = \frac{C_N}{\|\delta\|} \cdot \Lambda \cdot \mathbf{P}^{1/2} \cdot \frac{L - g(\mathbf{x}_0)}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}} \cdot \frac{\mathbf{P}^{1/2} \cdot \Lambda^T \cdot \delta}{\sqrt{\delta^T \cdot \Lambda \cdot \mathbf{P} \cdot \Lambda^T \cdot \delta}} = \frac{C_N}{\|\delta\|} \cdot \frac{L - g(\mathbf{x}_0)}{\delta^T \cdot \Sigma \cdot \delta} \cdot \Sigma \cdot \delta.$$

9.3.2 Gradient of Value at Risk

Here, the gradient

$$\nabla_\delta \mathsf{L}_q = \frac{z_q}{\sqrt{\delta^T \cdot \Sigma \cdot \delta}} \cdot \Sigma \cdot \delta$$

is derived in two ways.

Method 1: Term-by-term differentiation with respect to δ of the expression for $\mathsf{L}_q(\delta)$ gives us

$$\nabla_\delta \mathsf{L}_q(\delta) = \nabla_\delta\left(g(\mathbf{x}_0) + z_q \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right) = \nabla_\delta g(\mathbf{x}_0) + \nabla_\delta\left(z_q \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right)$$
$$= 0 + z_q \cdot \nabla_\delta\left(\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right) = z_q \cdot \frac{1}{2 \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \nabla_\delta\left(\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}\right)$$
$$= \frac{z_q}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}$$

Method 2: Apply the formula from 2.3.4 expressing $\nabla_\delta \mathsf{L}_q(\boldsymbol{\delta})$ as the ratio of gradients of Q:

$$\frac{\partial \mathsf{L}_q}{\partial \theta} = -\frac{\partial Q/\partial \theta}{\partial Q/\partial L} = -\frac{\frac{C_N}{\|\boldsymbol{\delta}\|} \cdot \frac{\mathsf{L}_q - g(\mathbf{x}_0)}{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}{-\frac{C_N}{\|\boldsymbol{\delta}\|}} = \frac{\mathsf{L}_q - g(\mathbf{x}_0)}{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta} = \frac{z_q}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}.$$

9.3.3 Gradient of Tail Value at Risk

Here, the gradient

$$\nabla_\delta \mathsf{T}_q = \frac{\phi(z_q)}{q} \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}$$

is derived in two ways.

Method 1: Term-by-term differentiation with respect to $\boldsymbol{\delta}$ of the expression for $\mathsf{T}_q(\boldsymbol{\delta})$ gives us

$$\nabla_\delta \mathsf{T}_q(\delta) = \nabla_\delta\left(g(\mathbf{x}_0) + \frac{\phi(z_q)}{q} \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right) = \nabla_\delta g(\mathbf{x}_0) + \nabla_\delta\left(\frac{\phi(z_q)}{q} \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right)$$
$$= 0 + \frac{\phi(z_q)}{q} \cdot \nabla_\delta\left(\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}\right) = \frac{\phi(z_q)}{q} \cdot \frac{1}{2 \cdot \sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} \nabla_\delta\left(\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}\right)$$
$$= \frac{\phi(z_q)}{q} \cdot \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}}$$

Method 2: We may use the conditional expectation theorem (sections 2.3.5, 9.2) in a manner parallel to the derivation of the gradient of exceedance probability above. As a consequence of linearity,

$$\nabla_\delta \mathsf{T}_q = E\left[\nabla_\delta g(\mathbf{X}, \boldsymbol{\theta}) \mid g > \mathsf{L}_q\right] = E\left[\mathbf{X} \mid \frac{\boldsymbol{\delta}^T \cdot \mathbf{X}}{\sqrt{\boldsymbol{\delta}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\delta}}} > z_q\right].$$

As before, transform variables to $\mathbf{X} = \boldsymbol{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \mathbf{Z}$ with spherical unit normal $\mathbf{Z}$, and then by a rotation to $\mathbf{Y}$ via $\mathbf{Z} = Y_1 \cdot \mathbf{u}^{<1>} + \sum_{j=2}^{N} Y_j \cdot \mathbf{u}^{<j>}$ where $\mathbf{u}^{<1>}$ is the "u" defined in section 9.3.1.

Note that the $\mathbf{u}^{<j>}$ are orthogonal to each other. The random variable $\mathbf{Y}$ is still spherical normal, being related to $\mathbf{Z}$ by a rotation. Therefore, one can rewrite

$$\nabla_{\delta} \mathrm{T}_q = \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot E\left[Y_1 \cdot \mathbf{u}^{<1>} + \sum_{j=2}^{N} Y_j \cdot \mathbf{u}^{<j>} \mid \mathbf{u}^{<1>T} \cdot \mathbf{Z} > z_q \right]$$
$$= \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \left(E\left[Y_1 \cdot \mathbf{u}^{<1>} \mid Y_1 > z_q \right] + E\left[\sum_{j=2}^{N} Y_j \cdot \mathbf{u}^{<j>} \mid Y_1 > z_q \right] \right)$$

Since components Y_j are independent of Y_1, the second expectation is zero, and therefore (recalling the ξ function from section 5.1):

$$\nabla_{\delta} \mathrm{T}_q = \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot E\left[Y_1 \mid Y_1 > z_q \right] \cdot \mathbf{u}^{<1>} = \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \left(z_q + E\left[Y_1 - z_q \mid Y_1 > z_q \right] \right) \cdot \mathbf{u}^{<1>}$$
$$= \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \left(z_q + \frac{1}{q} \cdot \xi(z_q, 0, 1) \right) \cdot \mathbf{u}^{<1>} = \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \left(z_q + \frac{1}{q} \cdot \left(\phi(z_q) - z_q \cdot \left(1 - \Phi(z_q) \right) \right) \right) \cdot \mathbf{u}^{<1>}$$
$$= \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \left(z_q + \frac{1}{q} \cdot \left(\phi(z_q) - z_q \cdot q \right) \right) \cdot \mathbf{u}^{<1>} = \mathbf{\Lambda} \cdot \mathbf{P}^{1/2} \cdot \frac{\phi(z_q)}{q} \frac{\mathbf{P}^{1/2} \cdot \mathbf{\Lambda}^T \cdot \boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T \cdot \mathbf{\Lambda} \cdot \mathbf{P} \cdot \mathbf{\Lambda}^T \cdot \boldsymbol{\delta}}}$$
$$= \frac{\phi(z_q)}{q} \frac{\mathbf{\Lambda} \cdot \mathbf{P} \cdot \mathbf{\Lambda}^T \cdot \boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T \cdot \mathbf{\Lambda} \cdot \mathbf{P} \cdot \mathbf{\Lambda}^T \cdot \boldsymbol{\delta}}}$$
$$= \frac{\phi(z_q)}{q} \frac{\mathbf{\Sigma} \cdot \boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T \cdot \mathbf{\Sigma} \cdot \boldsymbol{\delta}}}$$

10. Figures

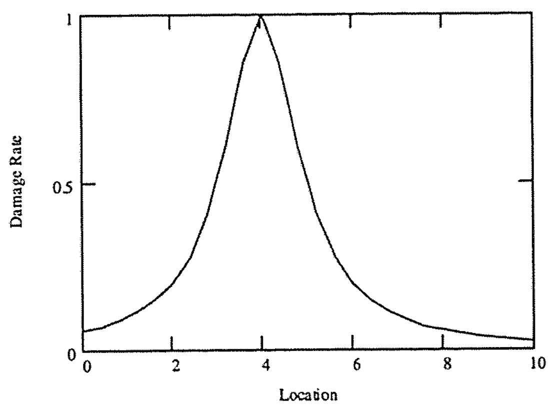


Figure 1: Profile of simulated hurricane damage, landfall at t=4, maximum loss rate r=1.

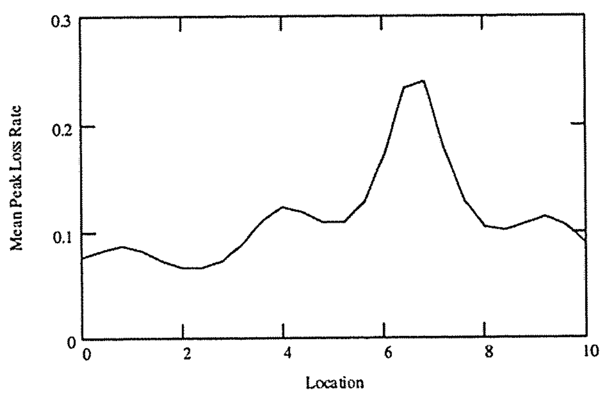


Figure 2: Conditional mean r as a function of location t.

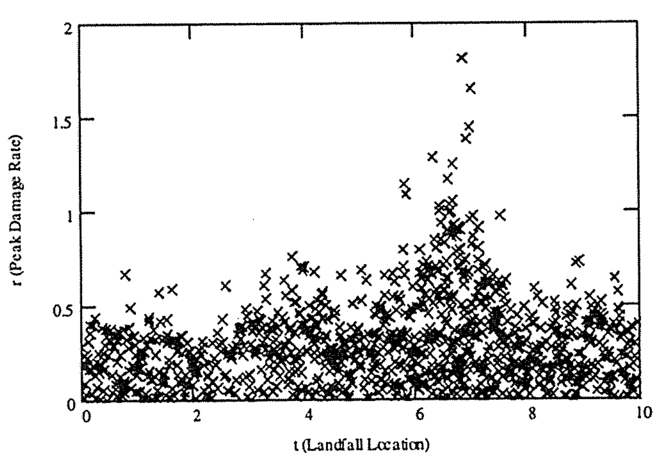


Figure 3: 1000 of the final 5002 events. Note: events with lower r values have higher associated weights.

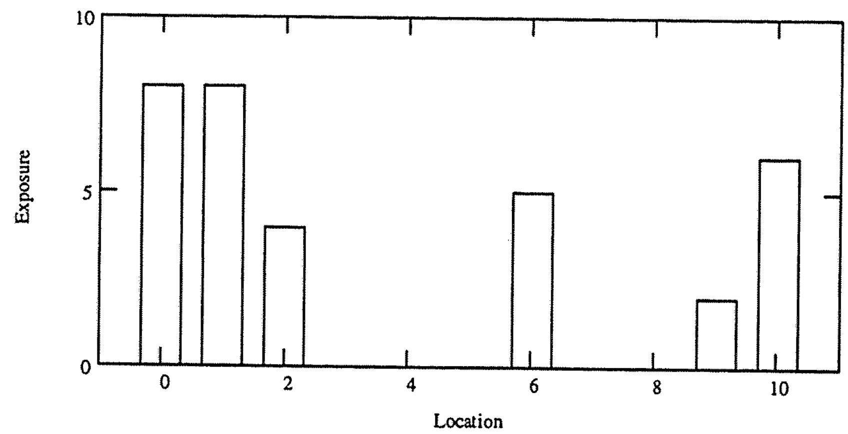


Figure 4: Insured exposures by location for simulation experiment

Location	Exposure
0	8
1	8
2	4
6	5
9	2
10	6

Table 1: Insured exposures by location for simulation experiment

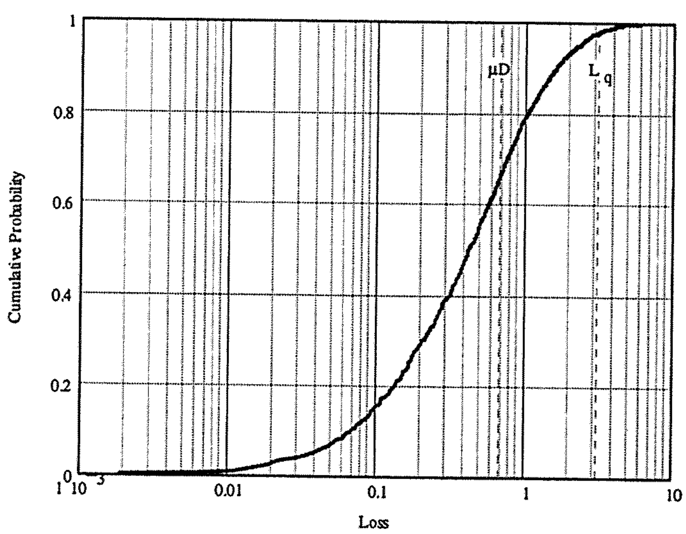


Figure 5a: Cumulative loss distribution for portfolio in figure 4

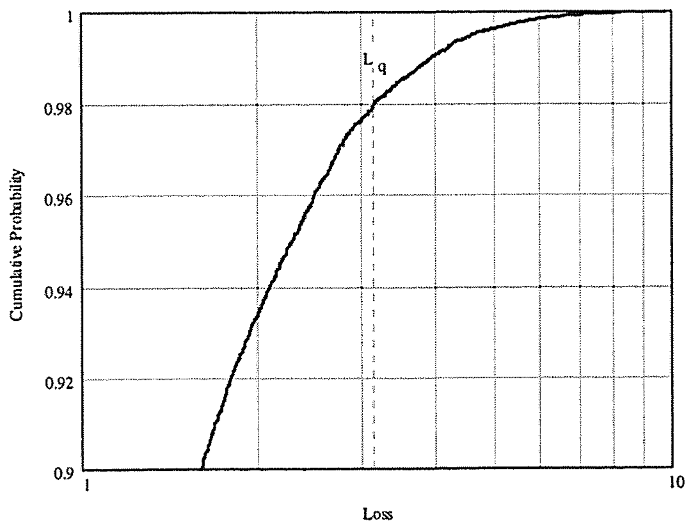


Figure 5b: Cumulative loss distribution for portfolio in figure 4 (closeup)

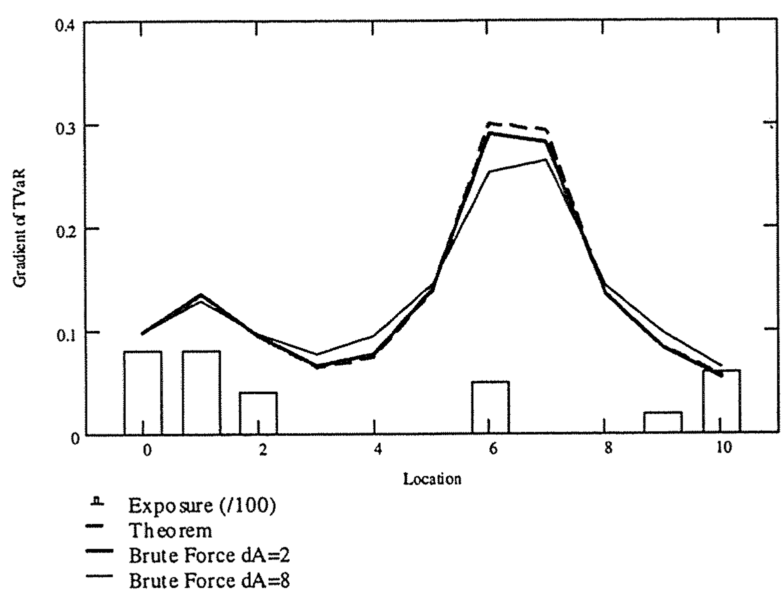


Figure 6: Comparison of methods for computing gradient of Tail Value at Risk with respect to exposures for ground up losses

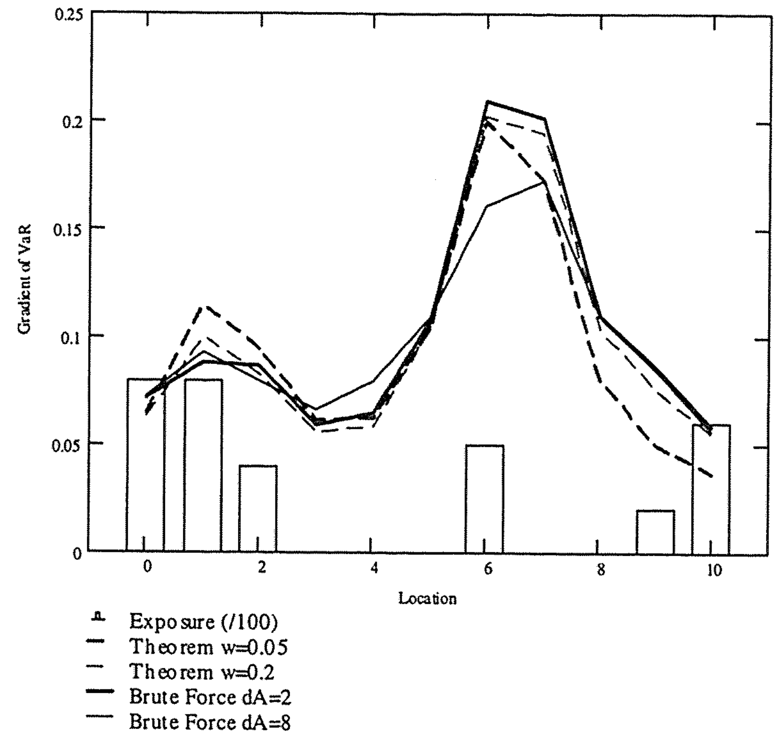


Figure 7: Comparison of methods for computing gradient of Value at Risk with respect to exposures for ground up losses

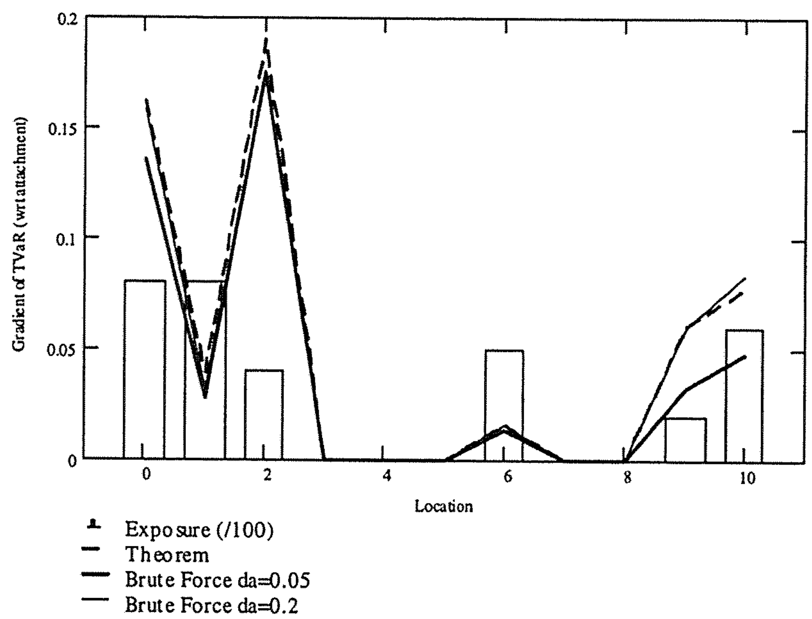


Figure 8: Comparison of methods for computing gradient of Tail Value at Risk with respect to attachment for losses net of treaty reinsurance

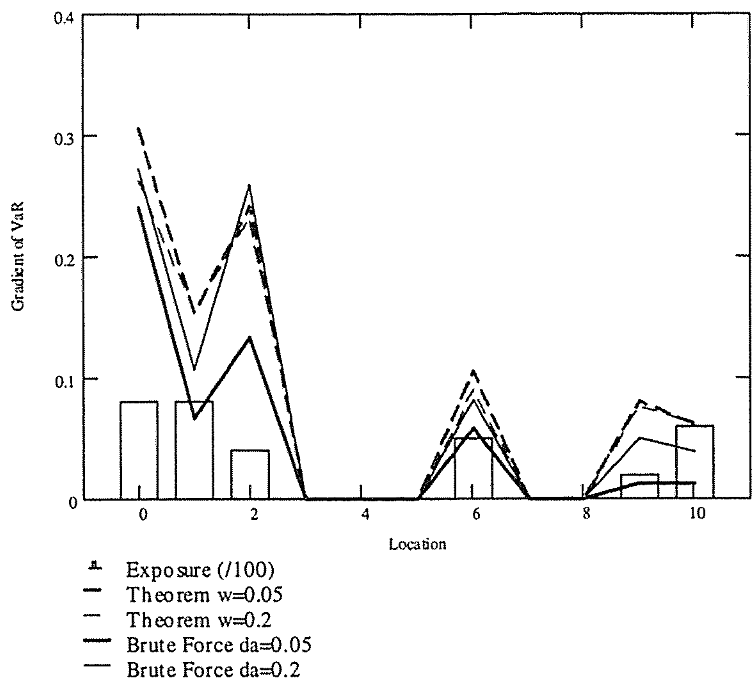


Figure 9: Comparison of methods for computing gradient of Value at Risk with respect to attachment for losses net of treaty reinsurance

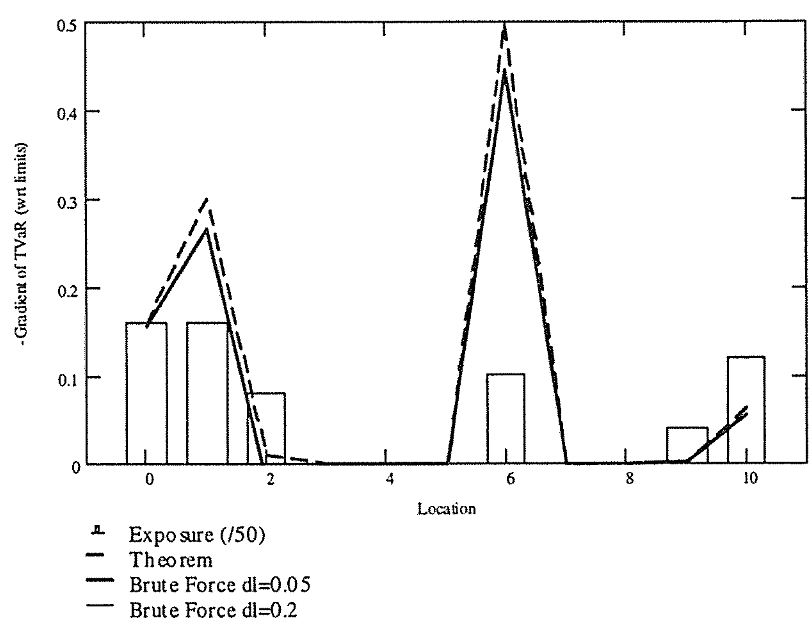


Figure 10: Comparison of methods for computing gradient of Tail Value at Risk with respect to limit for losses net of treaty reinsurance.

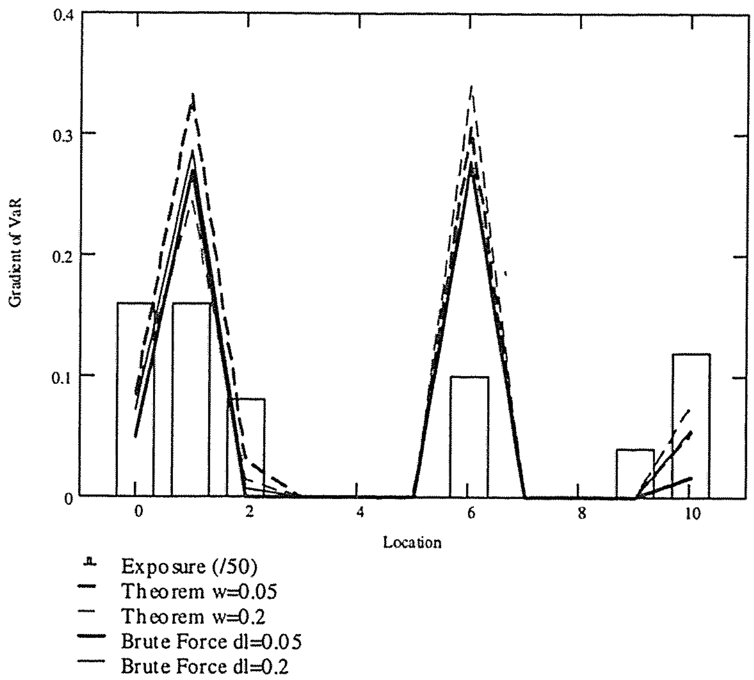


Figure 11: Comparison of methods for computing gradient of Value at Risk with respect to limit for losses net of treaty reinsurance.

	Exposure at Location											Att	Lim
	0	1	2	3	4	5	6	7	8	9	10		
New	8	8	4	0	0	0	5	0	0	2	6	2.3	1.2
Existing													
1	0	0	5	0	3	0	4	0	0	0	6	1.5	1
2	0	0	7	6	0	0	0	5	7	4	4	3	1
3	2	0	6	8	8	6	0	6	0	6	0	4	2
4	0	8	8	0	6	0	6	7	5	4	0	5	1
5	0	3	6	0	0	5	0	0	6	0	2	2	1
6	0	8	0	6	0	3	0	7	7	0	6	4	1
7	6	0	3	7	0	5	0	2	7	0	3	3	1
8	6	0	5	6	4	3	7	0	0	0	0	3	2
9	5	3	0	0	8	0	3	0	2	4	0	3	1
10	6	0	3	4	6	6	0	0	0	0	0	3	1
11	5	6	0	0	5	0	4	9	4	5	0	5	1
12	5	0	4	5	0	6	6	5	8	8	4	6	1
13	7	0	0	0	3	0	0	4	5	0	0	3	1
14	0	3	6	7	0	0	8	6	0	3	0	4	1
15	0	6	5	0	0	7	3	4	0	0	5	3	1
16	7	0	3	3	0	0	5	0	0	0	3	3	1
17	2	0	4	0	0	0	0	8	3	3	0	2	1
18	0	0	0	3	6	0	5	0	4	6	0	5	1
19	0	0	3	0	4	0	6	0	4	0	0	3	1
20	0	4	6	4	0	4	0	5	4	0	0	3	1
21	5	3	0	0	0	3	6	5	5	0	8	4	1
22	0	4	0	6	0	5	0	0	0	0	6	2	1
23	4	0	0	4	0	0	0	0	0	0	0	1	1
24	2	0	6	5	6	0	4	6	4	0	0	2	1
25	5	0	4	0	7	5	0	0	3	2	0	3	1

Table 2: Portfolio of reinsurance layers with proposed new contract.

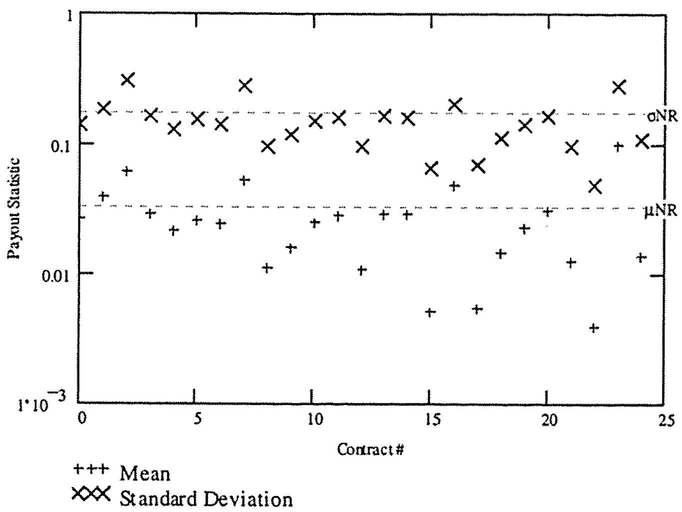


Figure 12: Payout statistics for new and existing contracts.

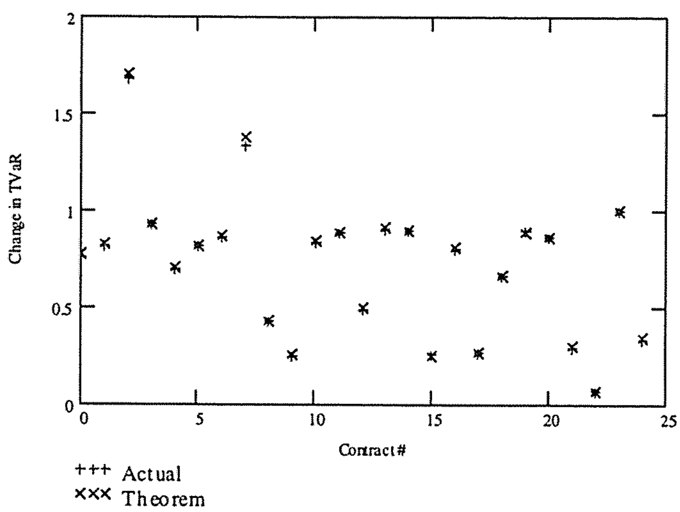


Figure 13: Comparison of actual and estimated change in TVaR upon cancelling each existing policy.

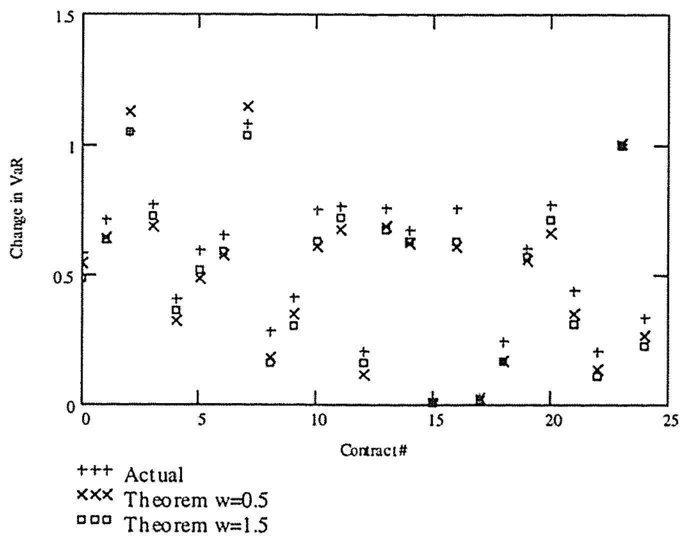


Figure 14: Comparison of actual and estimated change in VaR upon cancelling each existing policy.

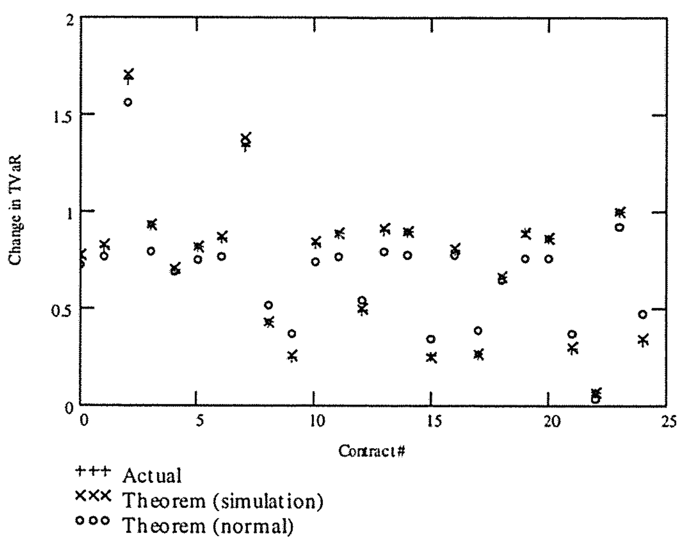


Figure 15: Comparison of actual and estimated change in TVaR upon cancelling each existing policy, including multivariate normal closed-form approximation.

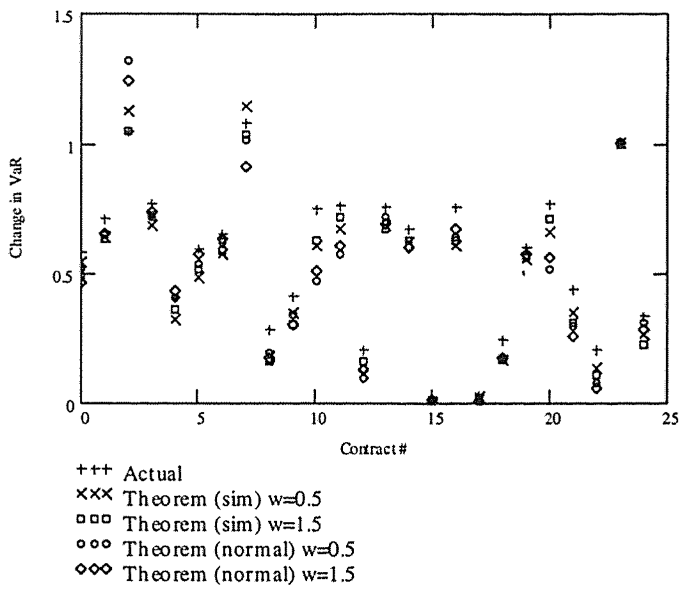


Figure 16: Comparison of actual and estimated change in VaR upon cancelling each existing policy, including multivariate normal closed-form approximation.

11. References

Abramowitz, M. and I. A. Stegun (1964) *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables,* Washington DC: National Bureau of Standards, U. S. Government Printing Office.

Acerbi, Carlo (2002) "Spectral measures of risk: A coherent representation of subjective risk aversion," *Journal of Banking & Finance* 26, pp. 1505–1518.

Anderberg, Michael R. (1973) *Cluster Analysis for Applications,* New York: Academic Press.

Andersson, Fredrik, Helmut Mausser, Dan Rosen and Stanislav Uryasev (2001) "Credit risk optimization with Conditional Value-at-Risk Criterion," *Mathematical Programming* Series B 89, 273-291.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999) "Coherent measures of risk," *Mathematical Finance* 9(3), 203-228.

Basel Committee on Banking Supervision (1996) "Amendment to the Capital Accord to Incorporate Market Risks," Basel, Switzerland: Bank for International Settlements.

Belubekian, Maya E. (2003) "Getting Results – and Exploiting Them," *Risk Transfer* Vol. 1 No. 5.

Berger, Adam J., John M. Mulvey, Kevin Nish and Robert Rush (1998) "A Portfolio Management System for Catastrophe Property Liabilities," Alexandria VA: Casualty Actuarial Society.

Britten-Jones, Mark and Stephen M. Schaefer (1999) "Non-Linear Value-at-Risk," *European Finance Review* 2: 161–187.

Daykin, C. D., T. Pentikäinen and M. Pesonen (1994) *Practical Risk Theory for Actuaries,* London: Chapman & Hall.

De Moivre, Abraham (1738) *Doctrine of Chances: A Method of Calculating the Probabilities of Events in Play.* Republished 1967 Irvington, Chelsea.

Fuglsbjerg, Brian (2000) "Variance Reduction Techniques for Monte Carlo Estimates of Value at Risk," Copenhagen, Denmark: SimCorp A/S.

Gaivoronski, Alexei A. and Georg Pflug (1999) "Finding Optimal Portfolios with Constraints on Value at Risk," *Proceedings of the III Stockholm Seminar on Risk Behavior and Risk Management,* January.

Gaivoronski, Alexei A. and Georg Pflug (2000) "Value at Risk in Portfolio Optimization: Properties and Computational Approach," Working paper, Department of Industrial Economics and Technology Management, NTNU (Norwegian University of Science and Technology).

Garman, Mark (1996) "Improving on VaR," *Risk* Vol. 9, No.5, pp. 61-63.

Gauss, C. F. (1809) *Theoria Motus Corporum Coelestium,* Hamburg: Perthes & Besser. (English translation by C. H. Davis, 1857, Boston: Little, Brown.)

Glasserman, Paul, Philip Heidelberger and Perwez Shahabuddin (2000) "Variance Reduction Techniques for Value-at-Risk with Heavy-Tailed Risk Factors," *Proceedings of the 2000 Winter Simulation Conference,* J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

Gourieroux, C., J. P. Laurent and O. Scaillet (2000) "Sensitivity Analysis of Values at Risk," *Econometric Society World Congress 2000 Contributed Papers.*

Hallerbach, Winfried G. (1999) "Decomposing Portfolio Value-at-Risk: A General Analysis," Technical Report, Erasmus University Rotterdam and Tinbergen Institute Graduate School of Economics.

Johnson, N. L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions Volume 1*, New York: Wiley-Interscience.

Jorion, Philippe (2000) *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill.

Klugman, Stuart A., Harry H. Panjer and Gordon E. Willmot (1998) *Loss Models: From Data to Decisions*, New York: Wiley-Interscience.

Kreps, R. E. (1990) "Reinsurer Risk Loads from Marginal Surplus Requirements," *Proceedings of the Casualty Actuarial Society* LXXX.

____ (1999) "Investment-Equivalent Reinsurance Pricing," in *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, O. Van Slyke, ed., Alexandria VA: Casualty Actuarial Society.

____ (2003) "Riskiness Leverage Models," Instrat Working Paper, New York: Guy Carpenter & Co., Inc.

Krokhmal, Pavlo, Jonas Palmquist, and Stanislav Uryasev (2001) "Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints" *Journal of Risk* Vol. 4 No. 2, Winter.

Laplace, Pierre S. (1783) "Sur les Naissances, les Mariages, et les Morts," *Mém. Acad. R. Sci. Paris*, pp. 693-702.

Larsen, Nicklas, Helmut Mausser and Stanislav Uryasev (2002) "Algorithms for Optimization of Value-at-Risk" in *Financial Engineering, e-Commerce and Supply Chain*, P. Pardalos and V.K. Tsitsiringos, eds., Kluwer Academic Publishers, 129-157.

Lemus Rodriguez, Gerardo José (1999) "Portfolio Optimization with Quantile-based Risk Measures," Massachusetts Institute of Technology, March.

Major, J. A. (2003) "Risk Contributions, Performance Measurement, and Co-measures," Instrat Working Paper, New York: Guy Carpenter & Co., Inc.

Major, J. A., M. Belubekian and S. Mathai (2001) "Portfolio Risk Management: Using the Gradient of Value-at-Risk," *Financing Risk & Reinsurance*, International Risk Management Institute, May.

Major, J. A. and R. E. Kreps (2002) "Catastrophe Risk Pricing in the Traditional Market," in *Alternative Risk Strategies*, M. Lane, ed., London: Risk Waters Group.

Manganelli, Simone, Vladimiro Ceci and Walter Vecchiato (2002) "Sensitivity Analysis of Volatility: A New Tool for Risk Management," European Central Bank Working Paper No. 194, November.

McGuinness, John S. (1969) "Is 'Probable Maximum Loss' (PML) a Useful Concept?" *Proceedings of the Casualty Actuarial Society* LVI.

Parkinson, C. Northcote (1958) *Parkinson's Law: The Pursuit of Progress*, London: John Murray.

Rockafellar, R. Tyrrell and Stanislav Uryasev (2000) "Optimization of Conditional Value-at-Risk" *Journal of Risk* Volume 2/Number 3, Spring.

Rouvinez, C. (1997) "Going Greek with VaR" *Risk* 10(2): 57-65.

Rubinstein, R. Y. (1981) *Simulation and the Monte Carlo Method*, New York: Wiley-Interscience.

Silverman , B. W. (1986) *Density Estimation for Statistics and Data Analysis*, CRC Press.

Studer, G. (1995) "Value At Risk and Maximum Loss Optimization," RiskLab Technical Report, ETH Zurich, December.

Studer, G. and H. J. Luthi (1996) "Quadratic Maximum Loss for Risk Measurement of Portfolios," RiskLab Technical Report, ETH Zurich, September.

Tasche, Dirk (1999) "Risk Contributions and Performance Measurement," Report of the Lehrstuhl für Mathematische Statistik, T.U. München.

Uryasev, S. (1995a) "Derivatives of Probability Functions and Integrals over Sets Given by Inequalities," *J. Computational and Applied Mathematics*, 56, 197-223.

____ (1995b) "Derivatives of Probability Functions and Some Applications," *Annals of Operations Research* 56, pp. 287-311.

____ (1999) "Derivatives of Probability and Integral Functions: General Theory and Examples," in *Encyclopedia of Optimization,* C.A. Floudas and P. M. Pardalos, eds., Kluwer Academic Publishers.

____ (2000a) "Introduction to the Theory of Probabilistic Functions and Percentiles (Value-at-Risk)," in *Probabilistic Constrained Optimization: Methodology and Applications*, S. P. Uryasev, ed., Kluwer Academic Publishers, pp. 1-25.

____ (2000b) "Conditional Value-at-Risk: Optimization Algorithms and Applications, *Financial Engineering News,* February.

Wang, Zheng (2002) "The Properties Of Incremental VaR In Monte Carlo Simulations, *The Journal of Risk Finance,* Spring.

Yamai, Yasuhiro and Toshinao Yoshiba (2002) "Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization," *Monetary And Economic Studies,* January.

Integrating Actuarial and Underwriting Disciplines to Improve Underwriting Outcomes

G. Chris Nyce, FCAS, MAAA

Integrating Actuarial and Underwriting Disciplines to Improve Underwriting Outcomes

Chris Nyce, FCAS, MAAA

About the Author:

Chris Nyce has held officer level positions in Actuarial roles for insurance companies in Ratemaking and Reserving. He also has held officer level underwriting positions in Global and National Accounts, Mid-Market Accounts, and Main Street Commercial Accounts. He is currently a Consulting Actuary who specializes in assisting companies in developing business improvement approaches that integrate Actuarial and Underwriting strategies. The topic and materials for this paper are drawn from a variety of actual implementations of integrated approaches using models such as the one discussed in this paper.

Abstract

This paper will present a practical method of integrating Actuarial and Underwriting objectives, with an emphasis on achieving correct rates, correct pricing, and correct reflection of individual risk characteristics in the final premium. It will discuss how implementing these in an insurance operation can improve synergy between the underwriting, actuarial, claims, and marketing functions of the company. The approach described relies on a meaningful segmentation of the book of business. By its nature the insurance business is an underwriting business, so integrating actuarial and underwriting

approaches should be a natural goal of the Actuaries. This paper contributes to the literature with powerful yet practical techniques practitioners can quickly employ.

Preamble:

Throughout the history of the CAS, the science of ratemaking, and the art of underwriting seemed to have diverged. Underwriters strive for the flexibility to enhance their ability to set a price that best reflects the individual risk characteristics insured, while Actuaries strive for the accuracy of the filed manual class rates to best reflect the risk potential in each class. However, filed manual rates do not always equate to the rates actually charged by Underwriters in the field, in part due to the subjective nature of many rating plans. In order to accomplish the objective of adequate rates, the gap between filed manual rates, and prices charged in the market must be reconciled.

The goals of Actuaries and Underwriters are the same, developing underwriting strategies that produce profits and growth. Despite the uniformity of objectives, Underwriters and Actuaries often seem at odds; each suspecting the other doesn't fully understand the insurance business. By its nature the insurance business is an underwriting business, so integrating actuarial and underwriting approaches should be a natural goal of the Actuaries.

This paper will present a practical method of integrating actuarial and underwriting objectives, with an emphasis on achieving correct rates, correct pricing, and correct reflection of individual risk characteristics in the final premium. The approach relies on meaningful and correct segmentation of the book of business. It will discuss how implementing these in an insurance operation can improve synergy between the underwriting, actuarial, claims, and marketing functions of the company. Approaches discussed in this paper most naturally are used in either the planning process, or in the rate review process. Other times when these approaches can be used would include new product introductions, in applying corrective actions to a book of business, or to further improve profitability in a book that is already performing well. Implementation of these approaches can also enhance the fairness of the prices and rates charged to policyholders.

In this paper, "rate" will denote the rate in the rating plan, while "price" will denote the final price charged to insured, as modified by subjective credits. Using this convention, it is clear that diligent Actuaries can set correct rates, but the final price charged in the market is in the hands of the Underwriters.

This paper presents an approach to reconcile the rate setting process, and the actual prices charged by desk Underwriters at the point the risk is accepted.

Common Misunderstandings:

To illustrate the misunderstandings that occur, many practicing Actuaries and Underwriters will recall variations of the following conversations while discussing insurance pricing.

ACT: Let me explain the commercial lines rate changes that we recommend filing in the state of Urbanois. We took an average of 10% matching your pricing goals. We identified the profitable classes, and took less of an increase or even a decrease in those classes. Unprofitable classes have been increased more than the 10% average. These changes will result in adequate rates in each class. With this change, you can achieve your overall pricing goals, and improve your class selection.

UW: Thank you for the rate change. The ten percent filed rate change matches our goals. We do need to amend the filing for expanded credits under our scheduled rating plan.

ACT: Why?

UW: Well, we noticed the rate changes vary by class. We will need to use the increased credits to bring each renewal risk in line with the 10% goal.

ACT: Is that all?

UW: No, we also need to use the expanded credits on the good classes. We will put tight guidelines in place to use maximum credits only on the good classes. Given the rate increase, we will also need the credits to attract new business.

ACT: Our meeting this morning may go longer than anticipated.

Misunderstandings often arise around the function of the filed manual rates. Underwriters will often use subjective rating to duplicate the function of the filed rates, to get the right rate by class. If the filed rates are correct by class, such actions by Underwriters will overshoot the needed changes. If the filed rates are not correct by class, perhaps due to filing limitations, subjective rating can still be used to obtain the correct price by class if appropriate discipline is used in the underwriting process.

Subjective rating is also often used to defeat the targeted changes of a rating plan. Striving for the perception of stability in the marketplace, Underwriters will smooth changes that vary by classification or other rating characteristic by using subjective pricing. In some cases, such as the introduction of simplified rating for commercial lines in ISO plans in the mid 1980's, most would agree some such approach might be reasonable. In more usual cases, with class rate changes correctly reflecting changes in loss experience by class, this can be counterproductive.

And on the other hand, Underwriters have an understanding of the business that is often missed by their actuarial colleagues. Without the careful evaluation of the individual merits of risks that the company accepts by the desk Underwriter, no company strategy can work. Yet often Underwriters view their job as rate setting, using rules of thumb to set price to exposure for business underwritten, and using flexibility within the rate plan to implement these rules, instead of being guided by the rate plan and using subjective rating to price based on quality of risk within class. Approaches shown in this paper will enable the Underwriters to focus on their critical role of individually underwriting risk characteristics, appropriately reflecting the quality of risk within class, while ensuring the achievement of the price and rate corrections implemented for the book of business overall in the rate review process.

ACT: Here is the indicated rate change. It shows we need an increase of 25%. When should we file?

UW: We have done a competitive rating study. This shows our rate structure is even with peers. At most we can take 5%.

ACT: How do you justify that?

UW: We'll re-underwrite the book.

ACT: How will the same Underwriters who underwrote this business to begin with achieve a timely additional 20% points improvement in results?

UW: I'll have to get back to you on that. Do you have any suggestions on how your rate study can help, other than to increase rates?

ACT: I'll have to get back to you on that also.

Actuaries often put too much reliance on the overall rate change in effecting improvement in the results. Underwriters understand more completely that the quality of the risks within the book and the proper use of subjective pricing to risk quality is as central to the performance of the overall book of business as is the appropriateness of the manual rate.

There is a worse outcome that could happen in the second conversation. The Actuary could agree to set an inadequate rate based on weak promises of re-underwriting and without an action plan to bridge the difference. Often Actuaries will cite the virtues of flexibility and good working relationships with the Underwriters when doing so. At certain times in the market, groups of companies file rates that undershoot the filed indications, hoping to bridge the difference with tighter underwriting. As a result, companies simply exchange business at lower rate levels. Without tight price monitoring on both new and renewal business, companies may not even be aware this is happening.

While it is clear such undisciplined actions adversely affect results of individual companies, on a macro level, these approaches can widen the extremes of soft market and hard market swings the industry, and companies, have experienced over the past decades.

Rates are often set by Actuaries in conjunction with home office underwriting or product management staff, and usually with local underwriting and marketing management. However, the rate change often is announced to the field Underwriters with a bulletin, without giving the Underwriters detailed knowledge on how to leverage and implement the new rate structure. Since the desk Underwriter sets the price for a risk, missing this link in the chain makes it impossible to achieve the intended rate actions.

The remainder of this paper will present a basic model, discuss approaches to designing a business strategy using the model, and finally present approaches to communicating and monitoring the successful implementation of that strategy.

Basic Approach:

The following key inputs needed to implement this approach are usually available in standard rate reviews:

- Overall profitability or rate change needed for the book of business
- Performance of the business by meaningful segment
- Market price competitiveness by segment

And to plan and achieve outcomes, the primary business drivers used in the model are:

- Retention by segment
- New business acquisition by segment
- Renewal price change by segment
- New business rate level by segment
- Underwriting approach by segment

The basic approach is to determine prospectively the performance of the book by meaningful segment, understanding the relationship of market price to the adequate price, and then plan pricing, business retention, and business acquisition to improve, reduce, or eliminate poorly performing segments of the book, while increasing the business in better performing classes.

Implementing this approach presents a number of questions to answer, and challenges to overcome:

- How to divide the book into meaningful segments.
- How to evaluate the performance of those segments prospectively.
- How to plan price, retention, and new business acquisition.
- How to predict the effects of those plans.
- How to effectively communicate the plan to field Underwriters.
- How to monitor to ensure the accomplishment of the plan.

Approaches to Segmentation:

There are advantages to a segmentation approach that uses easily available, objective characteristics to aggregate groups of risks. Loss ratio results and other business outcomes can then be driven and evaluated using these segments. Segmentation should focus on variables most important in determining prospective experience. For example, articulating a segmentation approach using "good risks" vs. "bad risks," and depending on Underwriters to determine this at a desk level, would be a poor way to segment a book. Such an approach is often called "re-underwriting". Underwriters are charged with evaluating good and bad risks within segments every day, and without some unusual circumstance, it is unlikely Underwriters will re-underwrite more effectively than they underwrite initially.

Often the best characteristics to use for segmentation are also characteristics used in rating, simply since the most important quantifiable characteristics tend to be in the rating plan. While it might seem reasonable that loss ratio results in breakdowns already accounted for in manual rate shouldn't be widely different, practical experience has shown that they often are. This is partly due to the operation of subjective rating to distort pricing as compared to manual rates. Other approaches to segmentation may use pricing characteristics used in rating, but at a greater level of refinement, such as using NAICS code for Commercial Auto rather than the vehicle use (secondary) classification codes. Or the most important characteristic might not be ratable, such as agency, or size of the risk.

While every situation is unique, it may be helpful to discuss some examples of how segmentation might be approached. NAICS code is a common and effective way to segment business. For a Commercial Property or Liability book, the occupancy class might be the most important. In a Personal Lines book, classification or geographic territory might be the most important. For D&O, the form of incorporation, or the nature of the business might be most appropriate. In Workers' Compensation, size of risk and classification could work, or for Commercial Motor, the vehicle use along with the size of the fleet could be most predictive. These are examples, but by no means the only, or even the best way to segment business.

Any degree of detail can be used to segment the business. One major characteristic can be used, or a cross between several. Credibility considerations should also be taken into account. A reasonable approach is that a great deal of detail will be evaluated in determining how the segmentation is done, but the final segmentation should be based only on the most meaningful level of detail. This makes it easier to communicate the plan to the many Underwriters that will be called on to implement the plan, remembering that if the plan fails at the point of individual risk implementation by Underwriters, it will fail overall.

Analysis of Market Prices:

Effective segmentation of the business requires good analysis of the market prices compared to prices that are required to achieve profitability. This analysis should be done in detail similar to the segments used in the model. Start by identifying the market leaders for the business evaluated, and be sure to obtain their rates. Often, the profitability by segment will be driven by the level of competition in that segment. The segments with the best historical experience will often be those for which competitors charge the most adequate rate. The worst performing segments will often have one or more lead competitors with inadequate prices. In general, the profit-maximizing rule will be to charge the higher of the available market price, or the adequate price for a segment.

Ways to Increase Predictability of Segmentation:

Segmentation is often based on past experience. Standard actuarial approaches are appropriate to evaluate the experience by segment, including putting premiums on current charged rate level, which often differs from current manual rate level. Past experience is a good starting point, but not sufficient in most cases to complete the evaluation. The consistency of the past data should also be examined. Consider questions such as;

- Is a segment showing poor results for each of the past five years, or is one very poor year influencing the analysis?
- Are similar segments showing the same results, and if so, can they be grouped together?
- Is industry experience or some other collaborating data available? Industry experience from statistical agencies can be a source of corroborating data.
- Does experience from competitors support the findings? Often competitors will file experience in enough detail, so that profitability by segment can be evaluated.
- Is the segment experience consistent? If a segmentation breakdown shows meaningful differences nationally, does this segmentation work in each region? If not, consider a different approach that does work consistently.
- Is the experience by segment consistent with the analysis of market price compared to adequate price? If the market price is inadequate, it is unlikely the segment can be grown profitably.

Finally, an environmental analysis should be done to evaluate how current events and trends could affect the analysis. Perhaps past experience doesn't reflect a recently discovered cause of action or cause of loss. Examples of this include the fairly sudden rise of mold claims, or construction defect claims that could influence future expectations for the performance of a segment of business that would not be reflected in the past experience. To accomplish this, it is critical to include Underwriters, Marketing professionals, and Claims professionals, in determining the segments. Their involvement in determining the segments will also help in obtaining the agreement and buy-in that is needed to implement the strategy, remembering that the strategy is implemented one risk at a time by Underwriters.

The better the confluence of corroborating data, market pricing factors, and environmental factors, the better the chances the segmentation will correctly predict the future that would occur if the model is not applied.

Note that this model can be used to effectively drive the results in the wrong direction if the segmentation is not done well. For example, one common pitfall is to treat segments which "everybody knows" are good business as good business without evaluating objectively. However, perhaps these segments may carry the least adequate market rates due to the general perception that these risks are "good", and the resulting heavier use of credits. Conversely, segments that "everybody knows" are bad business often carry market rates that are adequate due to the lack of aggressive competitors. The best approach is to be diligent in gathering several sources of information, and then evaluate it objectively. It is usually the surprises that represent the best opportunity.

Experience suggests that many books of business can be divided into three sections. The first produces all of the profit for the whole book. The second produces no profit, but does not produce losses either. The last produces operating losses, and these losses can be greater than all of the profits in the first section. Finding these segments represents a great opportunity, and Actuaries are very well prepared to take the lead in doing so. But finding the promising segmentation possibilities is not helpful if an effective action plan is not crafted to exploit them.

A further useful step is to put the results of the segmentation study into a form that can be easily communicated and implemented. Segments can be grouped into a small number of gradings. A simple approach is to separate the book of business into a small number of groups, perhaps 3 to 5, each with a meaningful share of the book of business, and assign a letter grade to each group, such as A, B, C, D or E. Note that these groups can contain any desired amount of detail. However, remembering that this segmentation will form the input into a plan implemented by the desk Underwriters, it is best to make sure the segmentation is clear enough to implement in that environment. That is, a large geographically diverse group will need to implement the findings, and communicate the plan effectively to the production force. Simplicity will also support the ability to consistently monitor and evaluate the key drivers, ensuring the plan is being followed.

The Core Model for Improving Underwriting Outcomes:

Core Model-Renewal Business:

Appropriate calculations to plan these business objectives are shown in Exhibit I. Once the book is segmented, and prospective loss ratios in the absence of the segmentation plan are known, it is a simple matter to determine for each segment what the renewal loss ratios will be after a renewal cycle with planned changes. An optimal outcome is achieved if the planned price changes achieve the same target loss ratio for each segment over one renewal cycle, i.e. the loss and ALAE ratios in line (12) of Exhibit I match. In some cases, a company may plan price changes to reach the target loss ratio in an extremely poor segment over two or more renewal cycles.

In the case where resulting loss ratios differ even after a renewal cycle, as in Exhibit I, prudent managers may plan to reduce the amount of business written in a poor segment. Even if price changes in a poor segment achieve the target loss ratio, it is reasonable to anticipate that corrective underwriting actions would lead to lower retentions in troubled segments, and a plan that shows that expectation will increase the chances that the corrective underwriting and pricing actions are taken. Individual risk underwriting actions will need to be most vigorous on segments with the greatest price increase. This will operate to reduce any anti-selection problems in the resulting loss ratio. For troubled segments, Underwriters need to determine what is causing the problem on a risk-by-risk basis. This is usually not an adequate substitute for pricing action, but instead must accompany it.

For all of these reasons, it is best to plan lower retentions in segments of the business where problems are being addressed.

Core Model-New Business:

The improvement actions taken on the renewal book can be lost, (and then some), if proper controls are not exercised on the new business being written. Proper pricing needs to be planned for new business. This can be done by filing and implementing correct manual rates. If a filing cannot be accomplished in a timely enough fashion, it may be appropriate to use targeted subjective pricing actions to accomplish a fair and adequate rate by segment. This can be done by setting targets for new business at a percentage of a benchmark price, such as a percentage of the filed rate. For segments with inadequate rates, this percentage is greater than 100%. Other benchmarks can be used, such as relationship to expiring rates on a similar book of business. However, if the filed rate plan is not the right base to set new business pricing, consider revising the rate plan.

It is appropriate to target new business volumes also. It is unlikely that a company can write large volumes of new business in segments which have performed poorly, or for which market prices are inadequate, and expect good outcomes. On the other hand, segments that have good performance represent the desired targets. Targets for new business can be expressed as percentages to existing business, or as absolute volumes. This needs to be monitored closely, so feedback can be given to desk Underwriters and Producers on any needed adjustments to their underwriting actions.

Similar comments apply to the individual risk underwriting approach for new business as with renewal. The most underwriting rigor is required for a segment which has been identified as troubled, especially if higher prices are being obtained, or prices charged are higher than the market price.

Retention:

Closely related to renewal pricing is retention. As expected, it is best to target higher retentions on segments with best prospects, and to measure frequently. Lower retention on problematic segments should not be achieved by pricing action alone. It is best to identify the problem areas, and selectively non-renew when there are good reasons to do so. Keep less of the problematic segments, but keep the best of them. Note that the example in Exhibit I has the poorest performing segment at an extremely low retention of 20%. If a segment with very adverse results can be identified, the optimal plan may have very low retention. It is not uncommon to find limited segments for which no retention at all may be optimal.

Planning and Predicting Outcomes:

Exhibit I shows how outcomes can be predicted based on this approach. The outcomes, loss ratio and premium volumes by segment, are determined by the business drivers identified above, pricing, retention, underwriting actions, and new business acquisition targets. The underwriting approach comes into play preventing slippage in predicted results due to anti-selection caused by the pricing actions. These underwriting actions may themselves provide improvement in the results, but it is prudent not to count on this improvement in most applications. Note that the new business loss ratio outcomes are usually not as favorable as similar renewal business loss ratios, often due to a culling of the book at first renewal for adverse underwriting characteristics that were discovered too late in the new business underwriting process to issue a cancellation notice until renewal.

This should be considered in the model. In the example in Exhibit I, new business was assumed to have a 5-point higher loss ratio after consideration of the planned rate changes.

The pricing, retention, and new business targets, and the resulting new and renewal mix of business determines the overall loss ratio outcome by segment. In the example, an overall rate change of 20.8% is achieved, but the outcome on the renewal book is better by 9.7 points of loss ratio when compared to improvement from applying the rate change without segmentation, as shown in Exhibit I. This improvement is achieved only to the extent that real and predictable differences in prospective loss ratios are discovered by the segmentation analysis, and the appropriate underwriting and pricing actions are planned and achieved. The degree of difference between loss ratios by segment shown in Exhibit I is typical of what can be found in a normal book of business.

Communicating and Achieving the Plan:

Once the model in Exhibit I is used to determine a plan, the achievement of the benefits depends on implementing and achieving the plan actions. Prices are determined at the desk of the Underwriters, so implementation at the underwriting desk is key.

Exhibit II shows an example of a simple chart showing the targets set for the key drivers (price, retention, new business appetite, and underwriting approach). This chart is designed to be a one page easy reference posted, perhaps, on the Underwriters' cube or on the office wall.

The degree of flexibility around the price targets should be made clear. The best practice would not be to require the same price change or achievement for every risk in a segment, but to expect to achieve the targeted price on average. Within each segment, the Underwriter differentiates the pricing based on the individual risk characteristics.

Along similar lines, the retention target might be the hardest target to hit precisely. Hitting the precise retention target is probably less important then the message to the Underwriter for poor segments, which is, underwrite for quality, and don't worry about losing a risk just due to proposing acceptable terms and price for coverage. On the other hand, in better segments, consider what needs to be done to keep the business, but still achieving targets and carefully underwriting the risk, of course. Monitoring should focus on making sure the relationship between retention in poor segments and target segments is appropriate.

Accompanying the chart in Exhibit II, Exhibit III shows a simple way to display the segment into which a particular risks fits. The example is one dimensional, and is based on occupancy classification. Depending on which segmentation is effective, this could be by classification, NAICS code, agent, territory, size of risk, other variables, or a combination. It could be by lines of business, or describe several lines for an account.

Note that the analysis involved in the segmentation exercise may be quite involved. However, boiling that analysis down to a simple implementation model is key to the ability to communicate and monitor the implementation. Exhibits II and III are designed to communicate a comprehensive approach into targets an Underwriter and Underwriting Assistants can ascertain at a glance.

Note that this exhibit shows separate segmentation for several lines of business. Some companies feel an account approach is best, and the segmentation could be done by NAICS code, for example, across all lines of business, if that is the way the company views its business. While this approach can work, it is worth noting that segmentation works better the more precise the segments, and precision at a line level gives better theoretical outcomes.

Integrating the Marketing Effort:

Success in the market is not necessarily achieved by doing everything various parties in the market would like the company to do, but instead by being clear on what the company is willing to do and does well. Another way to express this is "clearly and consistently articulate the underwriting appetite to the market." The guidance in Exhibit III also serves as an appetite guide for the production force. The clarity with which the company approaches the market will help profitability in a number of ways. First, the new business flow will be channeled into more favorable segments, as Producers understand the appetite, and begin to submit a higher proportion of the desired business. Second, the "hit ratio" (number of risks written as a ratio to number of risks submitted, or quoted) will be more favorable, as the company can be expected to offer more competitive terms to the more favorable segments, as driven by the targets in Exhibit II.

Higher hit ratios lead to lower expense ratios, less time quoting risks outside the company appetite, less time declining risks, more time for Underwriters to focus on underwriting risks in the desired segments, and more time for Underwriters to interact with the production force. All of these benefits from higher hit ratios can be obtained by channeling the submission flow into the segments the company desires to write.

Monitoring for Achievement:

All of the planning and communication discussed above is meaningless without a monitoring and feedback mechanism designed to gain corrections when key business drivers achieved differ from those planned. The more timely the feedback, and the more force behind achieving the targets, the better.

Monitoring the key business drivers falls into two areas, pricing, and production.

Developing the ability to monitor the true pricing in the marketplace is critical. While designing a price monitor is beyond the scope of this paper, numerous difficulties occur when designing an accurate price monitoring system. Just mentioning a few of the obstacles in monitoring renewal price changes can bring headaches to some Actuaries. These include mid term endorsements, cancel and rewrite transactions, optional coverage changes including limits and deductibles, and other details that must be accounted for in an effective price monitor.

Similar issues arise in designing new business price monitors. For the approach described in this paper to work, an effective price monitor must be built for new and renewal business, distributed to key players, and reviewed for target achievement. It also must be believable and trusted by all parties. Delivery at the level of the desk Underwriter is key, as these are the professionals that take the actions, are held accountable, and need to be able to effectively monitor prices on their own business.

Timeliness is also key. The best price monitor is one that delivers measurement including aggregated results, before renewal batches are released or new business becomes effective. This gives Underwriters the opportunity to change pricing prior to policy issuance if needed to achieve targets. If this is not possible, monthly monitors can still be effective. If the reporting frequency is quarterly or longer, the monitoring process becomes problematic, as the reporting is not timely enough to effectively respond and correct deviations from plan.

Price monitoring is a key discipline that many companies are rapidly adopting in response to the poor underwriting results of the soft market in the late 1990's. If implemented correctly, the price discipline can lead to better results overall, smoother results through the underwriting cycle, and more accurate reporting of results. If implemented poorly with the wrong targets, driving prices contrary to the market can lead to anti-selection and deterioration of results. Note that price monitors are as important in periods of declining prices as in periods of rising prices. Soft markets often result in market price declines that are more precipitous than company management is aware. To a great extent, this is caused by the aforementioned process of companies exchanging "new" business, with "new" business pricing that declines by greater amounts than any one company is aware.

Monitoring new business production and renewal retention is also best done monthly, or more frequently. As with pricing, deviations from plans need to be addressed quickly. Driving these measurements at the level of the Underwriter is important. For production measures, monitoring at the level of the accountable individual would suggest using agency level information.

Both pricing and production measurements must be made and summarized at the segment level for which the model is being driven and must also be at the level of the desk Underwriter who is accountable for achieving the targets to be most effective. Monitoring should also be summarized to levels of operating divisions (teams, branches, regions, divisions) for which managers are to be held accountable.

Summary:

The model above provides a way to instill discipline around pricing, underwriting, and marketing activities. Effective segmentation is key to obtaining the improved underwriting results that the model provides. Implementation can result in improvement in profits for a company using it, and can moderate fluctuations in results due to market swings. The clarity around the roles of Actuaries and Underwriters can result in both professionals focusing on the value added activities in which they excel. These approaches can lead to fair and adequate rates even if the rate filing process is problematic, as long as discretionary pricing is possible, and disciplined. Continual and timely monitoring of actual outcomes against the model targets is a requirement for effective use of the model.

Once the loss ratio benefits of this model are achieved, they are permanent, as long as the segmentation approach is updated periodically and an appropriate strategy implemented. This should be done at least annually, corresponding to one renewal cycle in most lines of business.

The ability of this model to improve company underwriting results requires clarity by management on strategic direction. Management cannot be unclear about risk appetite, and still implement these models. Care must also be taken to ensure that segmentation is correct. Both past performance, and market pricing are important considerations in how the segmentation is done. Ideally, the analysis of both results in similar conclusions. As with any pricing actions, implementation of the model should also be done with care to avoid adverse selection within segments. The model is a powerful tool to obtain improvements in results, but if implemented with segmentation done poorly or incorrectly, or put another way, if management has the wrong strategy, the model is also powerful in driving results in the wrong direction just as quickly.

Exhibit I

The Basic Model for Implementation

		A Classes	B Classes	C Classes	Total Book	
	Starting Position.					
(1)	Proportion of the book	29.0%	46.0%	25.0%	100.0%	From above
(2)	Loss and ALAE Ratio with no rate change	52.0%	69.0%	119.0%	76.6%	From historical experience
(3)	Rate change on renewals	10.0%	25.0%	50.0%	20.8%	Key variable to be driven.
(4)	Loss and ALAE ratio on new business	51.8%	59.2%	82.7%	56.1%	Assumed five points worse than expiring.
(5)	Net loss/exposure trend	1.5%	1.5%	1.5%	1.5%	Assume 4% loss, 2.5% exposure trend.
(6)	Assumed Unit count retention ratio	90.0%	70.0%	20.0%	63.3%	Key variable to be driven.
(7)	Resulting Premium Retention	99.0%	87.5%	30.0%	76.5%	Based on values in (6) and (3).
(8)	New business as percent of expiring premium	40.0%	35.0%	0.0%	27.7%	Key variable to be driven.
	Position after one year.					
(9)	Resulting premium growth	39.0%	22.5%	-70.0%	4.2%	Based on (7) and (8).
(10)	Resulting distribution of business	38.7%	54.1%	7.2%	100.0%	Based on (1) and (9).
(11)	Resulting loss and ALAE ratio on renewals	48.0%	56.0%	80.5%	54.7%	Based on (2), (3), and (5).
(12)	Resulting overall loss and ALAE ratio	49.1%	56.9%	80.5%	55.1%	Based on (4), (7), (8), (11).
(13)	Renewal experience if rate change alone were applied=					64.3%
(14)	Renewal Loss and ALAE ratio better due to segmentation strategy by					9.7%

(4)=[(2)+5%] / [(3)+1]
(7)=(6) X [(3)+1]
(9)=(7)+(8)-1
(10)={[(9)+1] X (1)} / [(9) for total book +1]
(11)={[(5)+1] X (2)} / [(3)+1]
(12)={[(11) X (7)] + [(4) X (8)]} / [(7) + (8)]
(13)=(2) x {[(5)+1] / [(3)+1]}
(14)=(13)-(11) for total book

Exhibit II

Sample of What a Chart to Accompany Appetite Guide for Desk Underwriter Could Look Like

(For each line of business, separate by region if desired)

Underwriters Targets for XXX Line of Business			
Class Group	A	B	C
Retention	90%	70%	20%
Renewal Price Change	+10%	+25%	+50%
New Business Appetite	Aggressively Seek	Open	None
Target % of Manual Rate	90%	100%	125%

Exhibit III

Sample Company Appetite Guide

For Use in Setting Targets, and for Communicating Appetite

Class Group	Line of Business I	Line of Business II	Line of Business III
Offices	A	A	A
Services	A	A	B
BuildingOwners	A	A	A
Light Manufacturing	A	B	C
Contracting	C	B	C
Wholesalers-Durable Goods	A	B	A
Office Condo	A	A	A
Residential Condos	B	A	A
Shopping Centers	A	A	A
Churches	C	B	B
Clubs	B	B	B
Hotels	B	C	C
Restaurants	B	C	C
Retail-not separately listed	B	B	B
Listed Retail Classes	C	B	A
Auto Services	C	C	B
Wholesalers-Non durable goods	C	C	C

The Cost of Conditional Risk Financing

Frederic F. Schnapp, ACAS, MAAA

Title

The Cost of Conditional Risk Financing

Contact Information

F. Schnapp
Director of Actuarial Analysis and Research
National Crop Insurance Services, Inc.
7201 W. 129th St., Suite 200
Overland Park, KS 66213

Telephone: (913) 685-2767
Fax: (913) 685-3080
Email: ffschnapp@yahoo.com

Abstract

This paper develops a risk pricing procedure by examining the role of capital in an insurance transaction. An insurance transaction differs from an investment in that an insurer uses capital at the time a claim is settled rather than when the policy is issued, and only if the damages exceed the premium for the exposure. The premium for the risk transfer, based on treating the use of the insurer's capital as a loan to the policyholder, is the amount required to ensure that the insurer's risk and return are in balance, with the expected loan payment representing the risk margin in the premium. The insurer's cost for providing these loans depends on the returns available on alternate investment opportunities. Risk diversification within a market segment is assumed to benefit the policyholders through reduced prices, while risk diversification across market segments is assumed to primarily benefit the insurer through a reduction in its risk. The specific form on the insurer's risk pricing function can be determined provided that the insurer operates under a capital preservation criterion in which its losses in one market segment are financed out of the profits earned in other market segments. The paper extends the model to consider expenses, federal income taxes, investment income, supply and demand, the competitive market price, and the time value of money.

The Cost of Conditional Risk Financing

Introduction

In "Pricing for Systematic Risk," the author examined the effect of risk diversification on price within a portfolio. This led to the development of a model for pricing insurance exposures based solely on their contribution to the systematic risk of the portfolio. Unlike financial pricing methods, the resulting prices were determined without regard to the insurer's capital, capital allocation techniques, risk adjusted rates of return, or the insurer's cost of capital. The model also placed no restrictions on the size of the portfolio. The portfolio might correspond to a single market segment for a single insurer or to the entire book of business for the insurance industry. In addition, the risk margin for each exposure was determined by allocating the risk margin for the portfolio based on the exposure's systematic risk. The model was not able to provide any information regarding the proper risk margin for the portfolio.

The purpose for this paper is to continue the examination of risk diversification and the insurer's price. Feldblum (1992) and D'Arcy and Dyer (1997) discuss a variety of financial methods that can be used to evaluate insurance risk margins. The Capital Asset Pricing Model has already been considered in "Pricing for Systematic Risk." Discounted cash flow methods, including net present value and internal rate of return models, are also frequently used for this purpose. One problem with discounted cash flow models is that the risk margin is based on the timing of the expected cash flows rather than on the uncertainty of the potential damages. As noted in Mango (2003), this type of approach evaluates the loading for risk "in an essentially deterministic framework." In addition, discounted cash flow models treat an insurance transaction as an investment decision. Capital is allocated to an insurance policy at the time the policy is issued, but unlike an actual investment, the capital is not actually spent. The reason this is misleading is that an insurer uses its capital at the time a claim is paid rather than at the time the policy is issued, and only if the damages exceed the premium for the policy. This differs from an investment in that the amount of capital required is not a fixed amount determined in advance. In order to provide a proper examination of the role of capital in support of an insurance transaction, the analysis needs to consider the gain or loss for each potential outcome. This approach is developed in the following section.

The Role of Capital in an Insurance Transaction

In order to examine the use of the insurer's capital in support of an insurance transaction, the following discussion will assume that all damages occur and are paid at time 0 and that the policy has no transaction expenses. Unless otherwise specified, the policyholder is assumed to purchase full coverage so that the indemnity payment equals the damages incurred. Each exposure will be considered in isolation from all other exposures so that the effect of risk diversification can be ignored. The notation P_X or $P(X)$ will be used to represent the insurer's premium for exposure X. Since the insurer requires an opportunity to earn a profit, the premium must be no less than the expected damages, $P_X >= E(X)$. In addition, the premium should not be so large that the insurer has no risk of incurring a loss, so that $P_X <= max(X)$. The insurer's use of its own capital depends on the actual outcome for the exposure. Whenever the damages x are less than the premium, the insurer would earn a profit of $P_X - x$, but whenever the damages exceed the premium, the insurer would need to contribute an amount $x - P_X$ of its own capital in order to settle the claim.

In order to illustrate the role of the insurer's capital in support of an insurance transaction, let X be an exposure having three outcomes of $0, $500, and $3000 with probabilities of .25, .50, and .25, respectively. Table 1 shows the insurer's results for each outcome based on the selection of a premium identical to the expected damages of $1000. The first four columns of Table 1 are self-explanatory. The column labeled "Return" represents the insurer's profit or loss for each outcome. Since the premium is equal to the expected damages of $1000, the expected return shown in the final row is $0. The final column, labeled "Deficit," shows how much capital the insurer contributes in order to settle the damages for each outcome.

Consider each of the outcomes for this transaction. If outcome A or B were to occur, the insurer would earn a profit of $1000 or $500, respectively. The insurer can use these profits in any manner it chooses, such as by adding these gains to its capital account, paying bonuses or stockholder dividends, or spending the gains in some other fashion. The only limitation on these funds is that they are not returned to the policyholder. The reason for this restriction is to prevent the policyholder from diversifying its risk over time. Otherwise, if the policyholder could offset the losses in one year with the profits in another year, its premium would be equal to its expected damages. As a result, the insurer would have no opportunity to earn a profit and insurance would not exist.

Table 1
Insurer Financial Outcomes Based on a Premium of $1000

Outcome	Premium	Damages	Probability	Return	Deficit
A	1000	0	.250	1000	0
B	1000	500	.500	500	0
C	1000	3000	.250	-2000	2000
Expected	1000	1000	1.000	0	500

Next, consider the insurer's results if outcome C occurs. In this situation, the insurer would contribute $2000 of its own capital in order to fund the deficit. In order to recover its loss on the transaction, the insurer could treat the $2000 as a loan that the policyholder would repay in future years. In order for the insurer to replenish its capital funds by the time of the next expected occurrence of outcome C, the loan would need to be repaid within four years in accordance with its annual probability of .25. Disregarding interest charges, the annual payment on the loan would be equal to the expected deficit of $500. In the subsequent year, the premium would be $1500, the sum of the expected damages of $1000 and the loan payment of $500. This amount would be paid each year until the next occurrence of outcome C or until the loan is fully repaid, whichever came first. Since more than one loan may be outstanding at the same time, the policyholder's premium may go up or down over time in response to the actual damages incurred. Since the insurer's funds are recovered in the years following the capital contribution, the procedure illustrated in Table 1 can be considered to determine a retroactive premium for the exposure.

From the insurer's perspective, the problem with the pricing procedure described in Table 1 is that it requires the insurer to recover its capital contributions from the policyholder in the years following the occurrence of outcome C. Since the policyholder can purchase the next year's policy from another provider, the insurer has no assurance of recovering its capital contribution. One way the insurer can address this problem by charging for the anticipated loan payment in advance of when the capital contribution is needed, rather than for the years following the occurrence of outcome C. Since the expected loan payment in Table 1 is $500, this suggests that the initial premium for the transaction should be $1500 rather than $1000. Table 2 evaluates the impact on the insurer's financial results using the revised premium of $1500. The revision to the premium changes the insurer's return and deficit for each outcome. For example, outcome C of Table 2 now shows that the insurer needs to contribute only $1500 of its capital in order to fund the deficit. This scenario also differs from the previous example in that the insurer expects to earn a profit of $500 in each year. Since the expected profit exceeds the loan payment (i.e., the expected deficit) of $375, the insurer expects to recover its capital contribution in three years rather than four. Consequently, the premium of $1500 can be considered to consist of the expected damages of $1000, the expected loan payment of $375, and an additional charge of $125.

Table 2
Insurer Financial Outcomes Based on a Premium of $1500

Outcome	Premium	Damages	Probability	Return	Deficit
A	1500	0	.250	1500	0
B	1500	500	.500	1000	0
C	1500	3000	.250	-1500	1500
Expected	1500	1000	1.000	500	375

Since the premium of $1500 in Table 2 overcharges the policyholder for the expected damages and the expected deficit, another scenario can be tested using an indicated premium of $1000 + $375, or $1375. This process can be continued for several more iterations until the process converges to a premium of $1400. At this point, the insurer's expected return (i.e., its expected profit) is equal to its expected deficit, as shown in Table 3.

Table 3
Insurer Financial Outcomes Based on a Premium of $1400

Outcome	Premium	Damages	Probability	Return	Deficit
A	1400	0	.250	1400	0
B	1400	500	.500	900	0
C	1400	3000	.250	-1600	1600
Expected	1400	1000	1.000	400	400

This process determines what might be described as the optimal retroactive premium for the exposure. Since the insurer charges the expected rather than the actual deficit, the premium of $1400 in Table 3 can also be interpreted as being the insurer's prospective premium for the transaction. At this price, the profits earned in years with favorable outcomes are just sufficient to pay for the insurer's capital contributions in the unprofitable years. Over a four-year period, the policy in Table 3 would be expected to generate a profit of $1600, which is just enough to fund the insurer's expected capital contribution of $1600 over the same period. The premium, which consists of the expected damages of $1000 and the expected loan payment of $400, corresponds to a loss cost multiplier of 1.40. Even though the insurer can still lose money over the short term, the premium should be satisfactory over the long term since the policyholder pays for the expected use of the insurer's capital.

The Insurer's Cost for Providing Capital

The pricing procedure developed in the previous section allowed the policyholder to borrow the insurer's capital without paying interest on the loan. It also considered a payment in the future as being equivalent to a payment in the present. The procedure can be made more realistic by including interest on the borrowed funds and by discounting the future payments to present value at the risk-free rate. For example, suppose that the insurer charges interest at the risk-free rate of 3%. Since the interest rate and the discount rate are identical, this has no effect on the premium of $1400 shown in Table 3. For outcome C, the insurer would contribute $1600 of its capital, which the policyholder would repay over four years with an annual payment of $430.44. Since the present value of this set of payments is $1600, the expected risk and return in Table 3 would be unchanged.

Under normal conditions, the interest rate charged on any loan should be expected to be greater than the risk-free rate. Suppose that the insurer obtains the necessary funds to provide the loan from its maturing investments. Since the insurer has the option of reinvesting the funds rather than providing the policyholder loan, the interest rate on the loan should account for the lost investment income potential on the insurer's capital. As a result, the interest rate being charged should be competitive with the returns available on other investments whose term is similar to the term of the loan. The risk of the various investment opportunities may also need to be taken into account in making this decision.

The example shown in Table 4 evaluates the required premium for the same exposure X considered in Table 3, but now charges 8% interest on the borrowed funds. The present value of the cash flows is based on a risk-free rate of 3%. In order for the transaction to be self-supporting over the long term in the same manner as the premium in Table 3, the expected return and the present value of the future loan payments should be in balance. Given the occurrence of outcome C, the insurer would contribute $1561.81 of its capital in order to settle the claim. At an 8% interest rate, the annual payment on the loan would be $471.54. Based on a risk-free rate of 3%, the present value of the four loan payments is $1752.77, a surcharge of roughly 12% over the amount of the loan. Since outcome C occurs with probability .25, the expected present value of these payments is $438.19, which matches the insurer's expected return. The premium of $1428.19 is an increase of $38.19 over the premium in Table 3.

Table 4
Cost of Conditional Risk Financing at an 8% Interest Rate

Outcome	Premium	Damages	Probability	Return	Deficit	Annual Payment	PV of Payments
A	1438.19	0	.250	1438.19	0	0	0
B	1438.19	500	.500	938.19	0	0	0
C	1438.19	3000	.250	-1561.81	1561.81	471.54	1752.77

Expected	1438.19	1000	1.000	438.19	390.45	---	438.19

This result can also be expressed algebraically. Let P represent an arbitrarily selected premium for the transaction. Given this premium, the insurer's return for outcome i is $P - x_i$ so that its expected return is $P - E(X)$. Whenever the damages x_i exceed the premium P, the insurer contributes capital of $x_i - P$ to settle the claim. Since the loan should be repaid by the time of the next expected occurrence of x_i, the term of the loan is $1/p_i$ years, where p_i is the probability of the outcome. A surcharge is included on each loan in order to recognize the insurer's cost of providing the funds. In lieu of an interest rate, the surcharge will be represented by a factor s_i that depends on the term of the loan. While this point is not essential for this discussion, it can be assumed that $s_i > s_j$ whenever $p_i < p_j$. Using this notation, the policyholder's cost of borrowing capital of $x_i - P$ is $(x_i - P)s_i$, so that its expected cost is $\sum_{xi>P} (x_i - P)s_i p_i$. Based on the procedure described above, the required premium P_X for the transaction is the value at which the premium is just sufficient to cover the expected damages plus the expected cost of the loan. That is, P_X is the solution to:

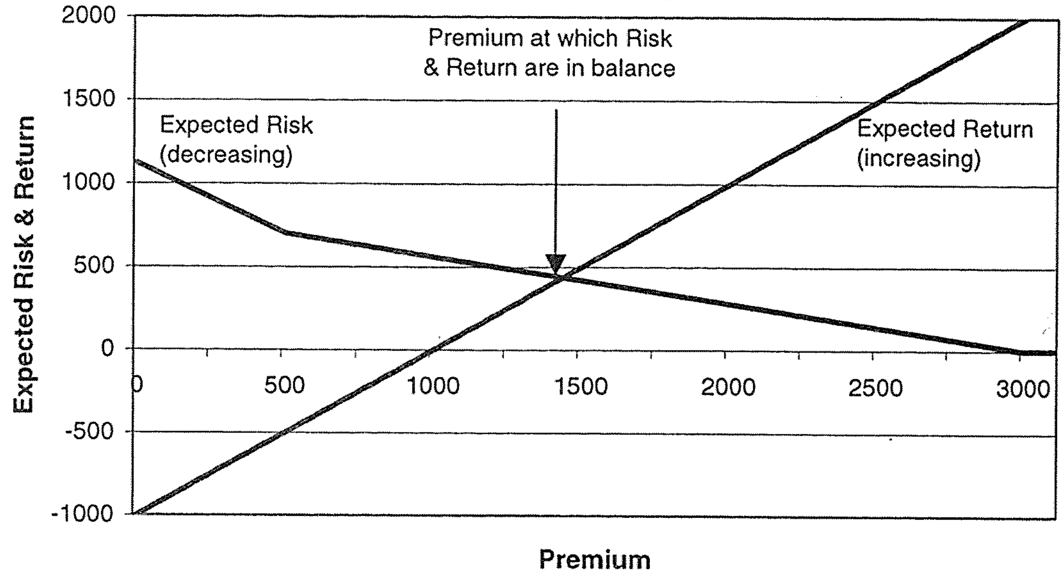
$$(1) \qquad P - E(X) = \sum_{xi>P} (x_i - P)s_i p_i$$

This result can be expressed more concisely by defining $(x_i - P)s_i$ as the risk for any adverse outcome x_i, that is, those outcomes for which $x_i > P$. With this definition, the right hand side of equation (1) represents the expected risk for the transaction while the left hand side represents the expected return. On this basis, P_X is the premium for which:

$$(2) \qquad \textit{Expected Risk} = \textit{Expected Return}$$

Figure 1 provides an illustration of the use of equation (1) to determine the premium for an exposure. The calculations are based on the three-outcome scenario in Table 4. Notice that the expected return is an increasing function while the expected risk is a decreasing function of P. At $P = 0$, the expected return is negative while the expected risk is positive. At $P = max(X)$, the expected return is positive while the expected risk is 0. Since the expected risk and the expected return are continuous functions of P, the two curves intersect at a single point. The intersection point determines the premium P_X where the expected risk and expected return are in balance.

Figure 1
Risk vs. Return for a Single Exposure



From the insurer's perspective, the disadvantage to equation (1) arises from its use of a different surcharge factor s_i for each probability p_i. The pricing procedure can be made more practical by replacing the insurer's surcharge factors s_i with a single surcharge factor α that represents its average cost of conditional risk financing. Since each s_i factor represents a surcharge, the minimum permissible value for α is 1. With this modification, the insurer's pricing equation for a discrete valued random variable becomes:

$$(3) \qquad P - E(X) = \alpha \sum_{xi>P} (x_i - P)p_i$$

The general pricing formula for any random variable X is:

$$(4) \qquad P - E(X) = \alpha \int_{x>P} (x - P)\, dF(x)$$

Equation (4) will be referred to as the risk pricing model in the remainder of this paper. The equation indicates that the model determines the risk margin based solely on the uncertainty of the damages for the exposure. Other risks, such as liquidity or reserving risks, have no bearing on the indicated price. This means that pricing risks, such as changes in rate adequacy across an entire insurance market pricing cycle, also have no effect on the price for the exposure. This prevents an insurer from being rewarded with a higher profit margin simply due to its earlier pricing decisions. The model also makes no differentiation between the profitability of stock and mutual companies. In addition, the model isn't concerned with the amount of capital held by the insurer or its target return on equity. Instead, capital is taken into account only through the amount of capital consumed for each outcome.

Self-Insurance

In the previous discussion, the insurance transaction focused exclusively on the insurer's point of view. By considering this transaction from the policyholder's perspective, the policyholder's self-insurance premium for the exposure can be determined. In place of a premium, the self-insurance program would have an out-of-pocket limit that would represent the maximum annual amount that the self-insurer would pay to reimburse any damages. The funds to pay for damages in excess of the out-of-pocket maximum would need to be obtained from other sources. The cost of these funds could differ depending on their source. For instance, relatives or friends of the self-insurer could provide the funds as a gift rather than as a loan. Another possibility is that they might loan the money to the self-insurer at below market rates. Funds could also be borrowed from banks or other commercial lenders at a higher interest rate. Since the self-insurer would want to prevent its debts from accumulating, each loan should be repaid by the time of the event's next expected occurrence in $1/p_i$ years. By viewing the policyholder as the insurer, this makes it possible to use equation (1) to determine the self-insurance price for the exposure.

Even if insurance is more expensive than a policyholder's self-insurance price, there may be reasons other than price for the policyholder to continue to purchase insurance. One of the benefits to self-insurance is that it avoids some the insurer's expense loadings, such as commissions and premium tax. An advantage to insurance is that it provides financing for the potential damages at a predetermined cost. For self-insurance, financing would be obtained after the damages occur. A self-insurer takes the risk that interest rates may be excessive and that financing may not be available at any price. If this prevents the damaged property from being repaired, it can have the effect of converting a partial loss into a total loss. A decision to self-insure would need to consider whether the benefits of self-insurance offset its disadvantages.

Risk Diversification and Price

Risk diversification has two effects on an insurer's price. As discussed in "Pricing for Systematic Risk," diversification reduces the price an insurer requires for a portfolio of exposures. In addition, it affects the insurer's price for every exposure within that portfolio. The paper also noted that a portfolio can be defined as a single market segment or as all market segments combined.

One way to provide a solution to the problem of insurance pricing is to assume that each insurer diversifies its risk over all market segments. If $W_1, \ldots, W_n$ represent the cumulative exposures for all n market segments, and $W_T = \sum W_i$ is the total risk exposure over all market segments, then the systematic risk pricing model can be used to determine the risk margin for each market segment. Under the assumption that the damages in distinct market segments are independent, the risk margin M_i for market segment W_i is $k\sigma_i^2$, where $k = M_T/\sigma_T^2$. Unfortunately, this solution has several shortcomings. First, insurance exposures are not actively traded in a secondary market, undermining the entire rationale for this approach. Second, this method does not resolve the problem of determining the appropriate reward associated with taking risk. Instead, it simply shifts the problem from the individual market segments W_i to the insurance market W_T as a whole. Third, market competition may permit insurers to price each market segment for its own risk. Due to the absence of a secondary market, the market pressures that would cause insurance prices to decrease in response to risk diversification across market segments may not exist.

There are several reasons that can be provided to support the view that insurance prices should reflect risk diversification within each market segment but not across market segments. For instance, in other industries, companies use diversification across market segments to reduce their risk rather than to reduce the prices they charge their customers. Diversification across markets would be an ineffective business strategy if all of the benefits accrued to the customers through lower prices rather than to the company. Similarly, if expansion across market segments reduced insurer's profit margins without improving their risk vs. return tradeoffs, insurers would have little incentive to expand into new markets. A second point to consider is the overall profitability of the insurance industry. According to an ISO study of industry results, 2001 was the first year in history that the property/casualty insurance industry experienced a net loss after taxes. The assumption that insurers retain the benefits of risk diversification across market segments provides a reason why the industry is so consistently profitable, especially in comparison to the airline, steel, and automotive industries, all of which are frequently lose money. Third, insurers are obligated by the long-term nature of their liabilities to have sufficient capital to settle all claims. If an insurer lost a considerable portion of its capital, it might need to obtain additional funds in the capital markets. However, investors would generally prefer to use their money to fund investments that earn a future profit rather than use it to settle an insurer's debts. Since it may be difficult for an insurer to raise additional capital, capital preservation should be an important consideration in its business operation. One approach the insurer can use to improve its ability to preserve capital is to retain the benefits of risk diversification across market segments.

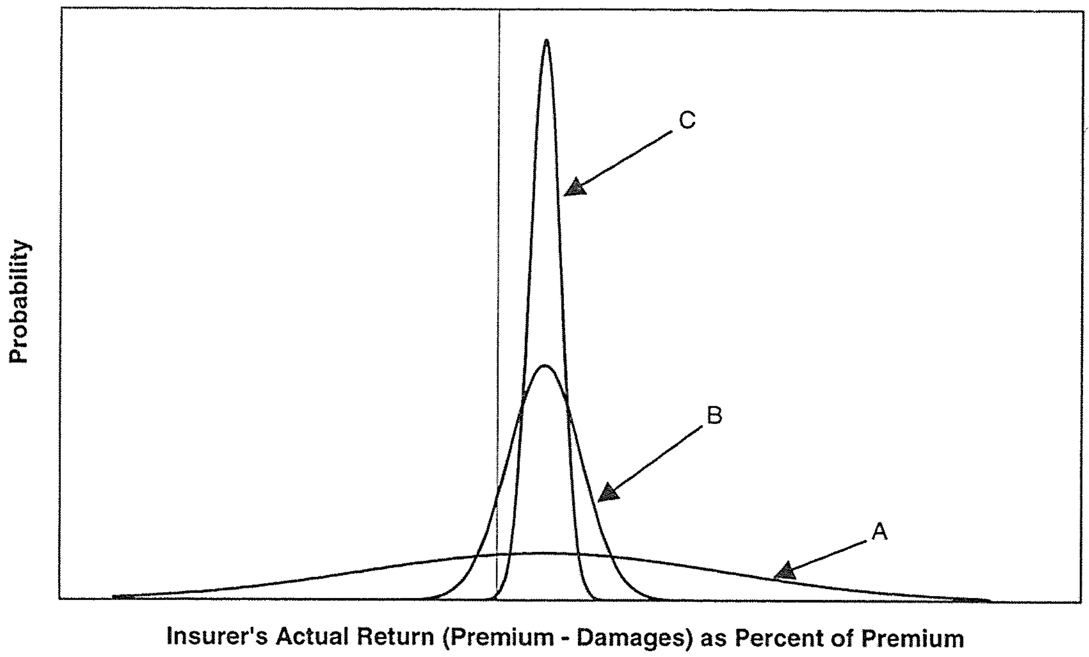
Based on the reasons provided above, this paper will adopt the assumption that an insurer's risk diversification within a market segment benefits its policyholders through reduced prices while diversification across market segments has little or no effect on the insurer's price. For the purpose of this discussion, a market segment will be considered to represent a group of exposures that are priced as a single entity, such as an insurer's Personal Automobile Liability book of business in a single state. For each market segment, the risk pricing model in (4) indicates that the insurer expects to be profitable but that it may experience a loss in any given year. Since the insurer would retain the benefits of risk diversification across market segments, this would reduce the insurer's probability of experiencing a loss over its entire portfolio. If the insurer is sufficiently well diversified and charges the prices indicated by the risk pricing model, it may be possible for it to be profitable over its entire portfolio virtually every year. Whether this is a reasonable characterization of the experience for large, well-diversified insurers will be left to others to address.

Figure 2 illustrates the effect of risk diversification across market segments on the insurer's financial results. The term "exposure" now refers to the experience of a single market segment. Each market segment is assumed to be priced for its own risk based on the risk pricing model in (4). For a single market segment, indicated by distribution A, the insurer has a significant probability of experiencing a loss. If it participates in five similar market segments, as indicated by distribution B, the insurer has the same expected profit as a percent of premium as for distribution A but with a much lower probability of experiencing a loss. If the insurer participates in 12 market segments, as indicated by distribution C, the insurer's expected profit as a percent of premium is still unchanged, but its probability of a loss becomes minimal. If the assumption that the insurer retains the benefit of risk diversification across market segments is valid, then the distribution becomes narrower, the profit becomes more certain, and the insurer's probability of a loss on its entire portfolio goes to zero as the number of market segments increases

One point that should be made about the results described above and shown in Figure 2 is that the description of the effect of risk diversification across a portfolio is more relevant to a large, well-diversified, primary insurer than it is to a reinsurer. Primary insurers tend to write relatively small independent exposures. Due to its small size, the outcome for any particular exposure would have little effect on the insurer's expected aggregate profit. In comparison, a reinsurer's portfolio may consist of exposures that are relatively large relative to the expected aggregate profit. Another issue is that reinsurance exposures may be more likely to be correlated with one another. Since this limits the reinsurer's ability to diversify risk across its portfolio, the reinsurer would be at greater risk of experiencing a loss on its entire portfolio. This suggests that a reinsurer's portfolio may be more consistent with distribution B rather than distribution C in Figure 2. Reinsurance and primary insurance also differ in that reinsurers tend to have long-term relationships with the primary carrier. Rather than attempting to earn a profit each year, a reinsurer may instead focus on the long-term profitability of the relationship. In this situation, the reinsurer could choose to recover its capital contributions retroactively in the years following a loss using the approach described in the discussion of Tables 1 and 2. Since a retroactive pricing strategy is inconsistent with the risk pricing model in (4), the insurer's aggregate profit and loss distribution would differ from those shown in Figure 2. Retroactive pricing may be a realistic approach for reinsurers to use since the damage distributions for their exposures may be uncertain. However, since the focus of this paper is on risk margins for prospective pricing, retroactive pricing in the reinsurance industry will not be given further consideration.

Before returning to the subject of risk diversification, a comparison between the reinsurance pricing approach described by Mango (2003) to the method described here may be worthwhile. In Mango's approach, "uncertainty is reflected between scenarios, not within them." His approach evaluates each scenario individually in order to determine the capital consumed for each outcome. His approach of evaluating risk by treating each outcome individually is consistent with the approach used in this paper. In addition, Mango's definition of the capital consumed for each outcome corresponds to the "deficit" as defined in Tables 1-4. Another similarity is that Mango determines a risk loading for each business segment rather than for the company as a whole. This is consistent with the earlier discussion of pricing each market segment for its own risk. One difference between the two methods is that Mango's surcharge on the capital consumed is based on the quantity consumed, while the pricing model in (1) bases the surcharge on the term of the loan. The risk pricing model in (4) uses a uniform surcharge factor in order to eliminate the need to identify the source of money used to fund the deficit. A second difference is that the risk pricing model in (4) considers the cost of the capital consumed to be the risk margin in the premium. Mango appears to develop the surcharge for capital consumption as a separate charge in addition to the risk margin in the premium. However, a careful examination of his approach suggests that this is not actually the case. The third and most important difference between Mango's approach and the pricing method described in this paper is that Mango focuses on the special problems of reinsurers. His approach develops the surcharge for capital consumption for each business segment individually, without examining the offset of profits and losses over the reinsurer's entire portfolio. The ability of an insurer to use diversification to reduce its risk and its capital consumption over their portfolio is the basic subject of this paper. In particular, this paper focuses on well-diversified insurers that are capable of successfully using diversification across market segments to eliminate almost all of their insurance risk, as indicated by distribution C in Figure 2.

Figure 2
Distribution of the Insurer's Returns



Risk diversification across market segments has a direct implication on the insurer's cost of providing conditional risk financing. In particular, it makes it possible to determine the value for α in equation (4). Consider an insurer that has successfully diversified its exposure over a large number of market segments. To illustrate, assume that the distribution of the insurer's outcomes for each market segment are similar to distribution A in Figure 2 but that its average results over its entire portfolio are similar to distribution C. Recall that the insurer uses its own capital only if the damages exceed the premium. Since the insurer is almost certainly profitable over its entire portfolio, the funds needed to pay for a deficit in the unprofitable market segments can be obtained from the gains in the profitable segments. The insurer would not be required to use its own capital to support its operation since it can obtain the required funds from the profits earned in the remaining market segments. In this sense, the policyholders would provide the funds needed to support the insurance operation without any use of the insurer's capital. Since the policyholder would no longer need to borrow the insurer's capital, the surcharge parameter α can be reduced provided that the insurer's portfolio distribution is still profitable. The insurer could continue to enter additional market segments until it reached the minimum value for α of 1. For a well-diversified insurer, this implies that the surcharge parameter α in equation (4) should be equal to 1.

The requirement that α must be equal to or greater than 1 can be related to existing research on risk pricing. A value for α of 1 implies that a well-diversified insurer should price each market segment so that the expected gain is twice its expected loss. This can be obtained by restating the risk pricing model in (4) for $\alpha = 1$ as:

$$(5) \qquad \int_{x<P} (P - x)\, dF(x) = 2\int_{x>P} (x - P)\, dF(x)$$

This result, which can be interpreted as *Expected Gain = 2*Expected Loss*, is consistent with Starmer (2000, p. 365), who states: "Benartzi and Thaler show that, assuming people are roughly twice as sensitive to small losses as to corresponding gains (which is broadly in line with experimental data relating to loss aversion), the observed equity premium in consistent with the hypothesis that investments are evaluated annually."

The conclusion that the surcharge factor α for a well-diversified insurer should be 1 is based on the assumption that the insurer has essentially eliminated its risk of experiencing a loss by diversifying over a large number of market segments. Other insurers may not be able to achieve this degree of financial security due to the limited size of their markets, statistical correlation between exposures, high expenses, or other reasons. In this situation, a value for α in excess of 1 may be needed in order for these insurers to achieve the same degree of financial security as a well-diversified insurer. For instance, suppose that the insurer's countrywide profit distribution based on a surcharge factor of $\alpha = 1$ is similar to distribution B in Figure 2. Since the insurer would have a significant probability of experiencing a loss, it would occasionally need to contribute its own capital to settle all of its claims. To correct this, the insurer can select a surcharge factor α in excess of 1 that shifts its aggregate profit distribution far enough to the right to ensure that it consistently earns a profit, but not so large that it becomes uncompetitive. In this sense, the value of α represents the insurer's degree of success in reducing its risk through diversification across market segments.

One limitation of the procedure described above is that since the tails of the distribution may extend to infinity, it may not always be possible to shift the distribution far enough to the right to ensure that the insurer would never need to use its own capital. One way to address this issue is for the insurer to purchase stop-loss reinsurance. The price for the reinsurance could be determined by applying the risk pricing model in (4) to the adverse outcomes in the insurer's aggregate profit distribution. Whether the net cost of the reinsurance is passed on to the policyholders would be a matter for the insurer to decide.

Comparison to Expected Utility Theory

The risk pricing model in (4) has been developed based on the role of capital in support of an insurance transaction. This formula can also be obtained more directly through the use of expected utility theory (EUT). One concern with the use of expected utility theory for the pricing of individual exposures is that it gives no recognition to the insurer's ability to reduce its risk through diversification within a market segment, as discussed in "Pricing for Systematic Risk." To avoid this concern, the EUT model will be used only to determine the insurer's price for a market segment as a whole. Given the insurer's price for the market segment, the prices for individual exposures within the market segment can be determined through the use of the systematic risk pricing model or any other method that provides an equitable allocation of the risk margin for the market segment to the individual policies that compose it.

Expected utility theory is a general decision making technique used by economists to evaluate an individual's choices under uncertainty. A detailed discussion of this subject can be found in Borch (1990) and Robison and Barry (1987). Expected utility theory is based on a series of axioms that describe an individual's preferences among simple and compound lotteries, as specified in Appendix A. The axioms are used to demonstrate the existence of a utility function that can be used to rank an individual's preferences. Since utility depends on preferences, each person may have a different utility function. An optimal decision is the one that maximizes an individual's expected utility.

The expected utility model is generally described in terms of a utility function $U(w)$, which is a continuous and increasing function of wealth, w. Utility is defined in terms of wealth rather than income in order to incorporate the individual's capital constraint into the evaluation of the optimal decision. Since the insurer is required to be risk averse, the utility function U is concave downward. The expected utility for an uncertain prospect (that is, a random variable) is determined from the utility for each outcome in combination with the probability distribution of the potential outcomes, i.e., $U(X) = \sum U(x_i)p_i$, where p_i is the probability corresponding to outcome x_i.

Traditionally, the insurer's certainty equivalent price for accepting a transaction for the uncertain damages X is defined to be the unique value $P(X)$ such that:

$$EU(w - X + P(X)) = U(w) \tag{6}$$

The interpretation of this formula is that $P(X)$ is the price that makes the insurer indifferent between the two options of accepting the exposure X for the premium $P(X)$ and not accepting the exposure. This formula compares the expected utility of the insurer's final but uncertain wealth of $w - X + P(X)$ to the utility of its initial wealth of w. The utility of initial wealth, $U(w)$, can be arbitrarily selected to be 0.

Expected utility theory pricing has the following basic properties:

$$P(c) = c \tag{7}$$

$$P(X + c) = P(X) + c \tag{8}$$

$$E(X) <= P(X) <= max(X) \tag{9}$$

$$\text{If } X(\omega) <= Y(\omega) \text{ for all outcomes } \omega, \text{ then } P(X) <= P(Y). \tag{10}$$

One important feature of the expected utility model in equation (6) is its dependence on the insurer's initial and final wealth. However, this relationship between utility and wealth is not required by the axioms of expected utility theory. By eliminating the reference to the insurer's wealth in equation (6), the expected utility theory model can be restated in terms of the effect of the insurance transaction on the insurer's income:

$$EU(P(X) - X) = 0 \tag{11}$$

The interpretation of this formula is that $P(X)$ is the price that makes the insurer indifferent between accepting the transaction for an uncertain gain or loss of $P(X) - X$ and not accepting the transaction.

At this point, assume that the insurer's price for a pro-rata portion of an exposure is the pro-rata price:

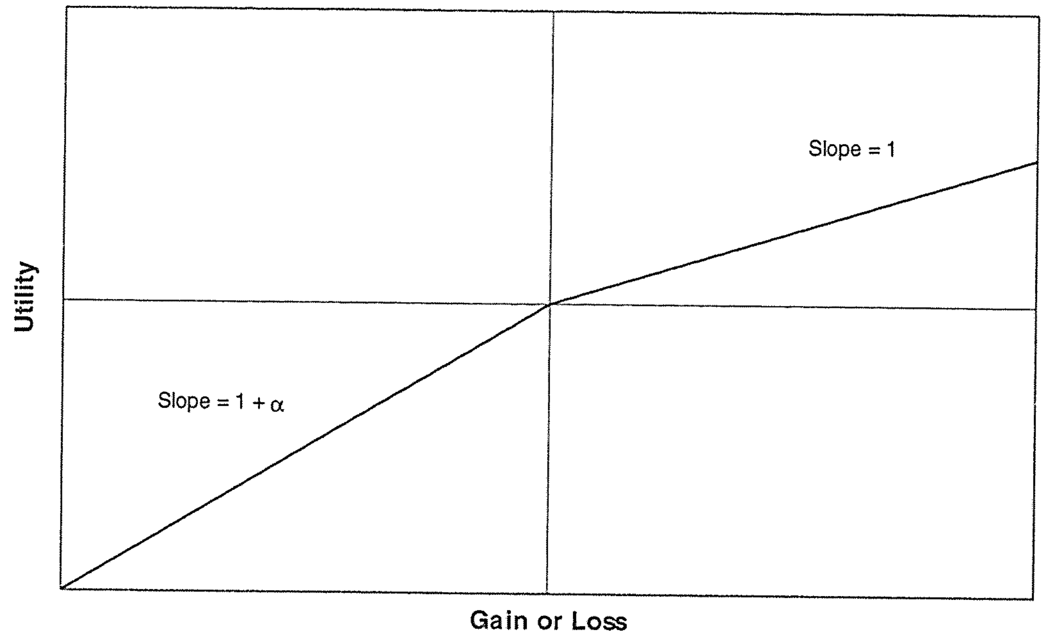
(12) $$P(aX) = aP(X) \text{ for } a >= 0$$

This property is essentially a statement about the effect of risk diversification on price. In "The Diversification Property," the pricing formula in equation (12) was demonstrated to be equivalent to the diversification property. This property states that $P(X + Y) <= P(X) + P(Y)$, with equality holding if and only if the two exposures are perfectly correlated with one another. The basis for the diversification property is that since diversification over imperfectly correlated exposures reduces the standard deviation, i.e., $\sigma_{X+Y} < \sigma_X + \sigma_Y$, it should also reduce the insurer's price. On the other hand, if the two exposures are perfectly correlated with one another, then no risk diversification occurs. Since this results in no reduction to the standard deviation, $\sigma_{X+Y} = \sigma_X + \sigma_Y$, there is no reduction in price.

It should be noted that the diversification property is not true for expected utility theory in general. For instance, if X is a Bernoulli random variable and the insurer's utility function $U(x)$ is $1 - e^{-0.2x}$, then it can be shown that $P(2X) > 2P(X)$. In addition, this particular utility function has the property that $P(X + Y) = P(X) + P(Y)$ whenever X and Y are independent. While this may appear to be an appealing characteristic, it implies that the insurer's financial condition (i.e., its utility) is not improved by insuring additional independent exposures. Moreover, the insurer's financial condition would become worse whenever the insurer's exposures are not all independent of one another. Since an insurer with this utility function obtains no benefit from risk diversification and may actually be in a worse financial condition after insuring a portfolio of exposures, the insurer has no incentive to provide coverage and may choose to withdraw from the insurance business instead.

The scalability property in (12) makes it possible to determine the general form of the insurer's utility function. Appendix B demonstrates that the utility function consists of two rays meeting at 0, as illustrated in Figure 3. The slope for the outcomes corresponding to gains can be arbitrarily selected to be 1, while the slope for the adverse outcomes is greater than or equal to 1. In this context, α represents the insurer's risk aversion parameter rather than its surcharge factor for the use of its own capital. Since the specific form of the utility function is known, it can be substituted into equation (11) in order to determine the insurer's certainty equivalent price. The resulting price function is identical to the risk pricing model in (4).

Figure 3
Shape of the Utility Function



The Standard Deviation Pricing Formula

The properties of expected utility theory in equations (7)-(10) and the scalability property in (12) make it possible to obtain a rough estimate of the required risk margin for each market segment. If the exposures in a market segment are independent, the central limit theorem states that the distribution of the average damages should be approximately normal. To make use of this result, consider a normally distributed exposure X with mean μ and standard deviation σ. Since X can be expressed as $\mu + \sigma Z$ where Z is a standard normal distribution, this gives $P(X) = \mu + \lambda\sigma$, where λ is defined as $P(Z)$. For an insurer with a risk aversion parameter of $\alpha = 1$, the value of λ as determined from equation (4) is approximately 0.3. Consequently, the risk margin for each market segment should be approximately 0.3σ.

As an application of this result, suppose that an insurer wants to determine the indicated rate change for a market segment. Assume that the trended current rate level loss ratios for the past five years are 0.70, 0.90, 0.80, 0.90, and 0.70, with a mean value of 0.80 and standard deviation of 0.10. Suppose that expenses are 35% of the premium and that the insurer expects to earn no investment income on the exposure. The total damages and expenses for the exposure can be expressed as $Y = X + e(Y)$, consisting of damages of X and expenses of $eP(Y)$. Applying (12), the premium for Y is $P(Y) = P(X)/(1-e) \cong (E(X) + 0.3\sigma)/(1-e)$. Using this result, the insurer's indicated rate change is (0.80 + (0.3)(.10))/(1-.35) – 1 = 27.7%. After applying the rate change, the indicated loss ratio and standard deviation become 0.627 and 0.078, respectively. Since the indicated risk margin at the revised rate level is 0.023 = (0.3)(.078) and the expense ratio is 0.35, the sum of the loss ratio, risk margin, and expense ratio is equal to 100% of the premium, as intended.

Prior to applying this result, it should be recognized that λ is approximately 0.3 when the insurer's risk aversion parameter α is 1. For other values of α, the value for λ would need to be recalculated. A second application of the standard deviation pricing formula is in testing the consistency of the insurer's risk margins across all of its market segments. If P, μ, and σ are for each market segment are known, and each market segment is approximately normally distributed, then the risk margins are consistent provided that $\lambda = (P - \mu)/\sigma$ is consistent across market segments.

Methods for Evaluating the Premium

In order to evaluate the certainty equivalent price for the uncertain damages, the risk pricing model in (4) must be solved for $P(X)$. This section reviews several techniques that can be used to determine the premium.

In simple cases, it may be possible to evaluate the integral directly. For example, suppose that the insurer's risk aversion parameter α is 1 and that X has the following distribution:

$$X = 1000 \qquad where\ Prob(X = 1000) = 0.50$$
$$X = 2000 \qquad where\ Prob(X = 2000) = 0.50$$

Since $1500 = $E(X) <= P(X) <=$ $2000, the expected risk component in (4) can be evaluated as $(\$2000 - P(X)) * 0.50$ while the expected return is $P(X) -$ $1500. Equating the two results and solving gives $P(X) =$ $1666.67. In this example, the premium consists of expected damages of $1500 plus a risk margin of $166.67.

The premium can also be obtained through a recursion process similar to that used in preparing Tables 1-3. Starting with an initial estimate P_1 of $P(X)$, a second estimate P_2 can be obtained by:

$$(13) \qquad \int_{x>P1} (x - P_1)\, dF(x) = P_2 - E(X)$$

By adding P_1 to both sides of the equation, this can also be expressed as $P_2 = P_1 + Expected\ Risk\ at\ P_1 - Expected\ Return\ at\ P_1$. For example, using the two outcome exposure introduced above, select P_1 as $E(X) =$ $1500. Since the integral in formula (13) can evaluated as ($2000 – $1500) * 0.50 = 250, this gives a second estimate of the premium of $P_2 =$ $1750. After several iterations, the premium converges to $1666.67.

A third method for evaluating the premium is to graph the expected risk and the expected return functions in equation (4), as shown in Figure 1. The intersection of the two curves determines the certainty equivalent price.

The fourth technique for determining the certainty equivalent price relies on the concept of synthetic probabilities. Synthetic probabilities represent a distortion of the actual probabilities to correspond to the risk of the exposure. Given the correct set of synthetic probabilities, the insurer's certainty equivalent price $P(X)$ can be evaluated as $E^*(X)$, where the expectation operator E^* is evaluated based on the use of the synthetic probabilities.

If the certainty equivalent premium $P(X)$ is known, the synthetic probabilities can be obtained by multiplying the actual probabilities by $(1 + \alpha)$ for those outcomes $x > P(X)$ and rescaling the entire set of probabilities to sum to 1. If $f(x_i)$ represents the density function or probability for outcome x_i, then the corresponding density function for the synthetic probability is $f^*(x_i)$, where:

$$(14) \qquad \begin{aligned} f^*(x_i) &= k(1 + \alpha)f(x_i) \quad \text{for } x_i > P(X) \\ f^*(x_i) &= kf(x_i) \qquad\qquad \text{for } x_i <= P(X) \end{aligned}$$

The value of k is selected to ensure that the total synthetic probability is 1. A detailed discussion of synthetic probabilities is provided in the discussion of the arbitrage theorem in Appendix C.

Synthetic probabilities can be especially effective in evaluating the premium for an exposure with a finite set of outcomes. The initial step in the evaluation is to arrange the damage outcomes in increasing order. The second step is to assume that the true premium that falls between a selected pair of outcomes x_i and x_{i+1}. Select a provisional premium P_1 between x_i and x_{i+1} and use the synthetic probabilities constructed in (14) to calculate $E^*(X)$. If $E^*(X)$ falls between x_i and x_{i+1}, then the assumption is true and the price $P(X)$ for the exposure is $E^*(X)$. If not, another pair of outcomes x_j and x_{j+1} can be tested. The new pair of outcomes should selected so that the new provision premium P_2 falls between the values for P_1 and $E^*(X)$ from the previous step. This process can be continued iteratively until the premium is determined.

As an application of this procedure, consider the exposure specified in Table 1. Since the expected damages for the exposure are $1000, the true premium must fall in the interval between outcomes B and C. For $\alpha = 1$, the synthetic probability procedure doubles the probability for outcome C to .500. After rescaling, the synthetic probabilities for outcomes A, B, and C become .200, .400, and .400, respectively. The expected damages using the synthetic probability distribution are $1400, which agrees with the premium determined in Table 3.

The Effect of Expenses on Price

The risk pricing procedure developed above can be easily modified to incorporate the insurer's expenses. Let X represents the insurer's experience for a single market segment. Suppose that in accepting the uncertain damages X in exchange for a premium of P, the insurer also incurs various expenses. These expenses represent the transaction costs arising out of issuing and providing service on the policies in the market segment. Any overhead expenses that cannot be directly attributed to the market segment will not be considered in this analysis. Let Y represent the uncertain loss adjustment expenses. The fixed expenses, i.e., those that are fixed dollar amounts independent of the premium, will be designated by f. Let the variable expenses such as commissions and premium taxes be expressed as a percentage g of the premium so that the amount in dollars is gP. Any federal income taxes owed will depend on the insurer's pre-tax profit or loss of $P - (x + y + f + gP)$. Assume that federal income taxes are based on a flat rate of t so that the federal income taxes owed are $(P - (x + y + f + gP))t$. Combining these results, the insurer accepts the uncertain outcome Z in exchange for a premium of P_Z where Z represents the total of the insurer's costs:

$$(15) \qquad Z = X + Y + f + gP_Z + (P_Z - (X + Y + f + gP_Z))t$$

Consider the simplest case, in which all expenses are fixed and taxes are zero. Setting Y, g, and t equal to 0 results in $Z = X + f$. Applying properties (7)-(10) to Z results in a premium of $P_Z = P(X) + f$. This indicates that the insurer's premium P_Z after taking expenses into consideration is equal to sum of the premium $P(X)$ based on no expenses plus the fixed expenses f. Following a similar approach, the premium for the exposure in equation (15) can be evaluated as:

$$P_Z = (P(X + Y) + f) / (1 - g) \qquad (16)$$

Notice that the federal income tax rate of t does not appear in the premium. This differs from the familiar ratemaking procedures described in Feldblum (1992) and Myers and Cohn (1987) that directly include federal taxes into the calculation of the required premium. The reason for this difference is that the Feldblum and Myers and Cohn models focus on obtaining a fair return for the investor, while the pricing model developed above looks at risk from the insurer's perspective. The interpretation of the result in (16) is that the taxation of underwriting gain and loss places the federal government in the role of a pro-rata reinsurer. The government can be considered to assume a portion, t, of the premiums, damages, and expense components of the insurance transaction, while the insurer retains the residual portion, $1 - t$, of the various components of the transaction. The insurer's risk and return components for the transaction are both reduced by the same factor, $1 - t$, resulting in no change to the insurer's certainty equivalent price.

The primary difference between income taxes and expenses is that expenses always reduce the insurer's income. Income taxes may be positive, negative, or zero, depending on the insurer's income. High expenses can make an otherwise profitable insurer unprofitable. Income taxes would reduce the insurer's income but would not make it unprofitable.

Before leaving this subject, two further points need to be clarified. First, the conclusion that federal taxes have no effect on the insurer's price applies only to a flat tax rate. Since the current federal income tax provisions differ from a true flat tax, a loading for income taxes may need to be included in the premium. However, this loading may be small enough to be ignored.

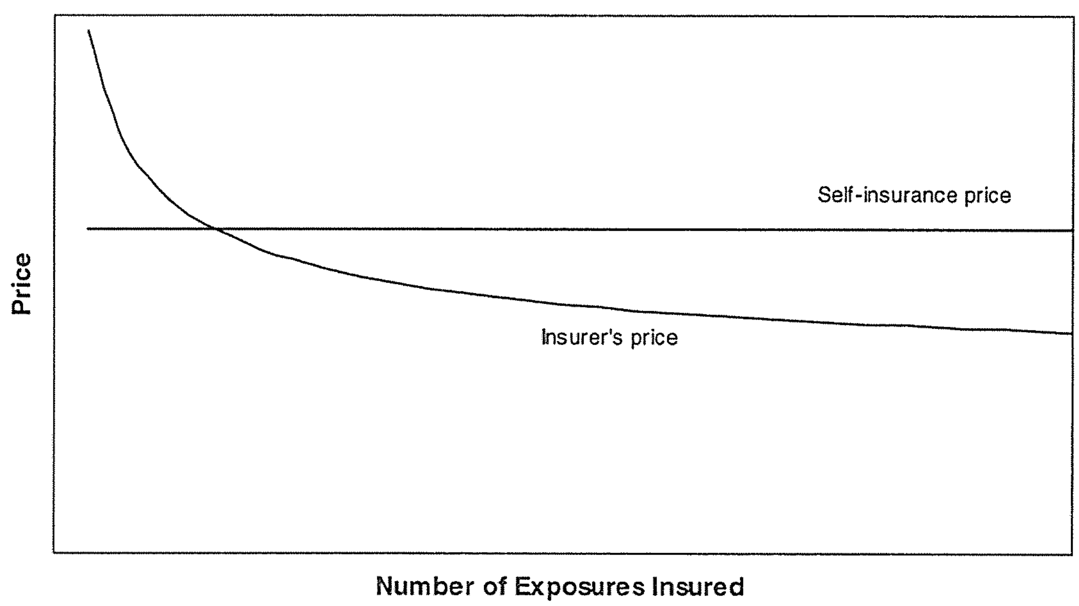
Second, even though federal income taxes have no effect on the insurer's price in a single market segment, they have a negative financial impact over the insurer entire operation. For example, suppose that distribution A in Figure 2 represents the insurer's experience for a single market segment. Since federal income taxes have no impact on the price the insurer charges, their primary effect is to narrow the width of the gain/loss distribution. This reduces the insurer's potential gain but also reduces its potential loss. Even though taxes have a neutral effect for a single market segment, this is not true over the insurer's entire operation. To illustrate this point, suppose that distribution C in Figure 2 represents the insurer's experience across its entire portfolio. Since the insurer has essentially no risk of experiencing a loss at this level of aggregation, federal income taxes provide no measurable benefit to the insurer in terms of reducing its risk of loss. Instead, taxes simply reduce the insurer's profits without providing an offsetting benefit.

The Mutually Acceptable Price

The following discussion returns to the subject of risk diversification within a market segment and its effect on the price for an exposure. In "Pricing for Systematic Risk," competitive pressures were described as compelling the insurer to reduce its prices to reflect its reduction in risk over the market segment. Even in the absence of competition, an insurer's price needs to be competitive with self-insurance. In order for a risk transfer to be acceptable to both parties to the transaction, the price offered by the insurer needs to be less than the policyholder's self-insurance price. Both prices can be determined using the risk pricing model in (4). However, the insurer's price includes its risk margin as well as an expense loading. If the expense loading is too large, the insurer's premium may exceed the policyholder's self-insurance price. In order to offer a competitive price, the insurer would need to take into account the risk diversification it achieves within the market segment in order to offer an acceptable price.

Suppose that the insurer has n independent and identically distributed exposures in market segment $Y = \Sigma X_i$, and assume that each policy has fixed transaction expenses of e. Using the approximation based on the standard deviation pricing formula, the insurer's required premium for the market segment as a whole is $P(Y + ne) = E(Y) + \lambda\sigma_Y + ne$, so that its price for an individual exposure X_i would be $P(X_i + e) = \mu + e + \lambda\sigma_X/\sqrt{n}$. As the number of exposures n increases, the insurer's price for each exposure would decrease. Provided that the transaction expenses e are not excessive (i.e., $\mu + e$ is less than the self-insurance price) and n is sufficiently large, the insurer's price would be less than the self-insurance premium as illustrated in Figure 4. In this situation, any premium between the insurer's price and the self-insurance price would be a mutually acceptable price for both participants in the risk transfer. However, if the insurer's expense loading e is excessive or the exposures are perfectly correlated with one another, a mutually acceptable price for the transaction may not exist.

Figure 4
Comparison of Certainty Equivalent Prices



The Effect of Deductibles on the Mutually Acceptable Price

The previous discussion found that a mutually acceptable price for a transaction exists provided that the insurer's expense loading is not excessive and that the insurer is able to diversify its risk over the market segment. Since the risk margin diminishes as the number of exposures increases, the existence of a mutually acceptable price hinges on the insurer's expense loading. The most convenient measure of an insurer's expense loading is its loss cost multiplier. For example, the exposure in Table 3 has a loss cost multiplier of 1.40 ($1400/$1000). Since this is large enough to accommodate a reasonable expense loading, a mutually acceptable price for this exposure may exist. For another exposure, the loss cost multiplier may be so small that the insurer would not be able to recover its expenses. For example, if the self-insurer's loss cost multiplier for another exposure is only 1.20 and the insurer's loss cost multiplier (representing the expense loading only) is 1.30, a mutually acceptable price for the risk transfer will not exist regardless of the number of exposures insured.

One technique an insurer can use to ensure the existence of a mutually acceptable price for an exposure is to offer a policy with a large deductible. For example, if the insurer offered a deductible of $500 on the exposure in Table 3, the loss cost multiplier for the excess coverage would rise to 1.60. Presumably, the self-insurer's loss cost multiplier for the excess coverage would be even higher. If the insurer's expense loading is 0.40 of the expected damages and its risk margin can be reduced to 0.20 through risk diversification, the insurer's actual premium would be based on a loss cost multiplier of 1.60, ensuring that a mutually acceptable price exists.

In determining the loss cost multiplier of 1.60 for the excess coverage, the insurer made the assumption that the policyholder evaluates the acceptability of the excess coverage independently of the decision to self-insure the damages below the deductible. More properly, the insurer also needs to consider whether the policyholder would prefer to self-insure both portions of the exposure. Based on the diversification property, $P(X + Y) <= P(X) + P(Y)$ for any two exposures X and Y. If X represents the retained exposure and Y the excess exposure, then the policyholder's self-insurance premium $P_S(X + Y)$ for the total exposure is less than the sum of its prices for the individual exposures, $P_S(X) + P_S(Y)$. In order for the policyholder to prefer to purchase the excess coverage for a premium of $P_I(Y)$ over self-insuring the total exposure, the combined premium $P_S(X) + P_I(Y)$ should be less than $P_S(X + Y)$ so that $P_I(Y) < P_S(X + Y) - P_S(X)$. Since the self-insurance premium $P_S(X)$ for the retained portion of the exposure in Table 3 is $428.57, the insurer's premium $P_I(Y)$ for the excess coverage must be no greater than $971.43. This limits the insurer's loss cost multiplier for the excess coverage to 1.55 rather than the 1.60 value determined above.

The Competitive Market Price

The policyholder's decision to purchase insurance or to self-insure an exposure is based on finding a mutually acceptable price for the transaction. By applying this analysis to every exposure, the supply and demand functions for the market segment can be determined.

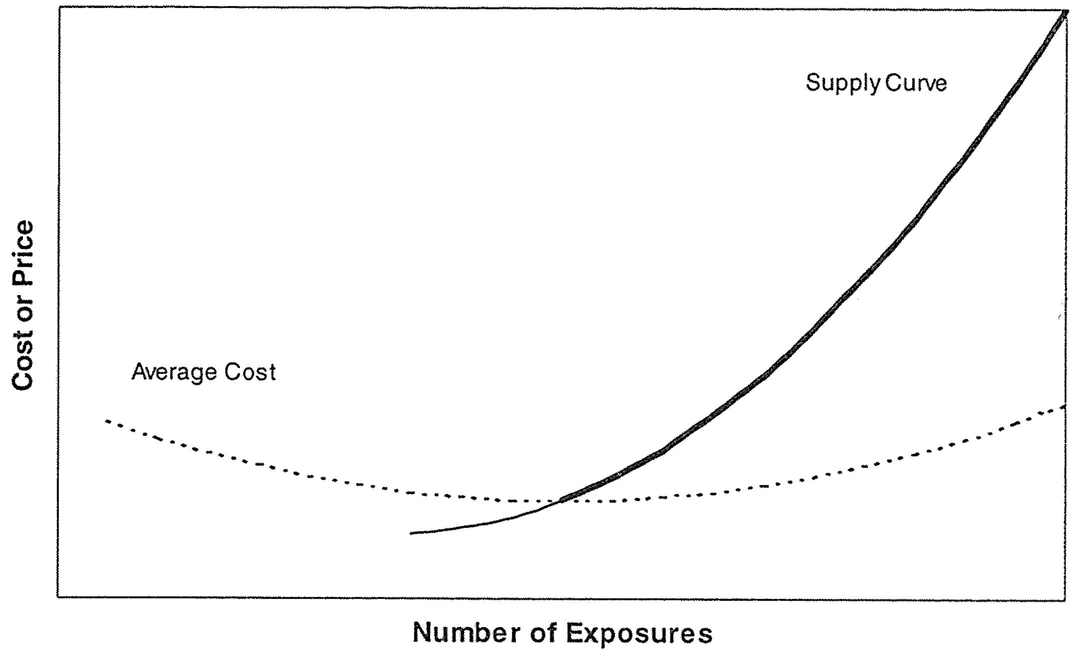
The first step in the analysis is to construct a demand curve for insurance. The demand curve represents the collective interest in purchasing insurance at a given price P. For any individual, the demand for insurance can be represented by a function whose value is 1 for any market price

P below the self-insurance price and 0 for any market price above that point. The market demand curve is the cumulative demand for insurance at each price, as measured by the total number of policies that would be purchased at a given price. As such, the market demand curve is simply the sum of the demand curves for each exposure.

The next step is to determine the supply curve for each insurer under conditions of perfect competition. These conditions require that each insurer have no influence over the market price and that exposures having identical risk characteristics are charged the same price. According to Thompson and Formby (1993), the supply curve for a particular company is based on its cost function. For an insurer, the total cost for a given number of exposures consists of the total expected damages, the insurer's fixed overhead expenses, the transaction expenses, and the insurer's cost of risk, where the cost of risk is the risk margin required by the risk pricing model in (4). The total cost can also be expressed as the sum of the insurer's certainty equivalent price for the market segment and the insurer's fixed overhead expenses. The insurer's total cost exceeds the insurer's breakeven cost for the market segment by the amount of the insurer's profit margin.

Once the insurer's total cost for the market segment is known, its average cost curve can be determined. This analysis will be limited to a discussion of the insurer's costs over the short term so that the insurer's staffing and other overhead expenses can be considered to be fixed costs. Due to risk diversification within the market segment and to a portion of the costs being fixed, the insurer's average cost curve should decrease as the number of exposures increases, as indicated in Figure 4. However, this overlooks the possibility that exposures differ in their risk characteristics. Due to imperfect information, rapid expansion by an insurer is likely to result in a lower quality book of business than its existing portfolio, thereby increasing the insurer's loss costs. As the number of exposures increases, the insurer will eventually reach a point of diminishing returns, i.e., a point where its average costs begin to increase. Consequently, the insurer's short-run average cost curve will be assumed to be "U" shaped as shown in Figure 5.

Figure 5
Insurer's Supply Curve



Differentiating the formula *Total Profit = Total Revenue – Total Cost* shows that the insurer maximizes its profit at the point where its marginal cost is equal to its marginal revenue. Since the market is assumed to be perfectly competitive, the insurer's marginal revenue is identical to the market price, where the market price can be any value along the vertical axis in Figure 5. If the insurer's average cost curve is "U" shaped as in Figure 5, the insurer's marginal cost curve will intersects the average cost curve at its minimum and remain above the average cost curve for all greater quantities sold. Since the insurer maximizes its profits by insuring the number of exposures indicated by its marginal cost curve, this means that the insurer's supply curve coincides with its marginal cost curve. The supply curve extends below the average cost curve until it reaches the insurer's average variable cost curve. At the intersection of the supply curve and the average cost curve, the insurer recovers all of its costs and earns a normal profit. For any market price above this point, the insurer would earn profits in excess of the normal profit, as measured by the vertical distance between the supply curve and the average cost curve. If the market price is below the intersection of the supply curve and the average cost curve, the insurer would earn less than its normal profit or may be unprofitable. If inadequate profits or losses persist over the long term, the insurer can withdraw from this market segment in order to pursue other market segments in which it can obtain a more adequate return.

The final step in the analysis of the competitive market price is to compare supply and demand for the market as a whole. Given the supply curve for each insurer, the market supply curve is the sum of the number of exposures that each insurer is willing to insure at each price. In a perfectly competitive market, the market price is the price at which the market supply and market demand curves intersect. At this price, the return earned by each insurer would differ due to the differences in their average costs.

The Time Value of Money

Up to this point, the discussion has been limited to exposures with uncertain damages whose actual outcome is discovered immediately after the risk transfer has been completed. The model will now be extended to damages in which the outcome is realized at a future time.

Suppose that all of the potential outcomes for X are realized at a future time t instead of time 0. The certainty equivalent price $P_t(X)$ for the risk transfer can be evaluated at time t using the risk pricing model in equation (4). Since this is a fixed amount, it can be discounted to time 0 at the appropriate risk-free rate, v_t, so that the price at time 0 to transfer the uncertain damages is:

$$P_t(X)v_t \tag{17}$$

The certainty equivalent price can also be evaluated using a second approach, by discounting each outcome for X to time 0 at the risk-free rate. Consider the random variable Y, defined as the value of the uncertain outcome X discounted to time 0 at the risk-free rate:

$$Y = Xv_t \tag{18}$$

Under this approach, the outcomes for X are now considered to occur at time 0. The certainty equivalent price for Y at time 0 is:

$$P_0(Y) = P_0(Xv_t) = P_0(X)v_t \tag{19}$$

The prices in (17) and (19) are identical provided that $P_t(X) = P_0(X)$, i.e., that the insurer's prices are consistent over time. For this to be true, the insurer's risk aversion parameter α must be stable over time. Since the risk aversion parameter α for a well-diversified insurer is 1, either of the two methods for determining the present value for an exposure with uncertain future outcomes will produce the same price.

In order to extend the result developed above to exposures with multiple payments made over time, let each x_i for $i = 1$ to m represent a single outcome for the uncertain damages X. Each outcome x_i represents a stream of cash flows $x_{i,j}$ with $j = 1$ to n, where j indicates the point in time at which the payment is made, so that:

$$x_i = (x_{i,1}, x_{i,2}, x_{i,3}, \ldots, x_{i,n}) \text{ for } i = 1 \text{ to } m \tag{20}$$

For each outcome x_i, the individual payments $x_{i,j}$ can be discounted to present value at the appropriate risk-free rate v_j. Replacing each outcome x_i with its present value $y_i = \Sigma x_{i,j}v_j$, the certainty equivalent value for a stream of payments X will be defined as $P(Y)$. This method meets two basic requirements. First, it produces the correct price for an exposure whose cash flows are certain, and second, the method is consistent with the pricing requirement in (19) for an exposure whose uncertain outcomes are paid at time t. Appendix C provides another perspective on this result by demonstrating the consistency of the risk pricing model with the arbitrage theorem.

Conclusion

The procedures commonly employed by actuaries to determine the appropriate risk margin in the insurer's rate are derived from financial analysis methods that evaluate risk from the perspective of the investor. This paper describes an alternate approach that determines risk margins based on the risk of the exposure, taking into account the insurer's ability to diversify its risk within and across market segments.

The pricing procedure is based on two concepts. The first is the role of capital in an insurance transaction. Capital is needed only when the damages for an exposure exceed the premium. The risk pricing model in (4) determines the prospective price for an exposure based on the requirement that the policyholder repay the insurer's capital contribution as if it were a loan. The pricing formula is consistent with an expected utility theory model based on a utility function consisting of two rays meeting at the origin. The price for a transaction is determined by finding the point at which the risk and return are in balance. Based on this approach, federal income taxes should have only a very limited or no effect on the insurer's price. The model also provides a means for determining the price for an exposure with future cash flows by discounting each cash flow to present value at the appropriate risk-free rate.

The second concept is that risk diversification affects price. The insurer is able to diversify its risk both within and across market segments. The paper assumed that risk diversification within market segments benefits the policyholders by reducing the insurer's price. Since insurers need to conserve capital, risk diversification across market segments has been assumed to have no effect on the insurer's price. For a sufficiently well diversified insurer, any losses in one market segment are covered by the gains in other market segments so that the insurer would rarely need to use its own capital to support its insurance operation. A less well-diversified insurer can improve its ability to conserve capital by increasing its risk aversion parameter α. The ability of policyholders to self-insure places a limitation on the insurer's ability to α to increase its risk margin. By explicitly recognizing the effect of risk diversification on price within a market segment, it was shown that a mutually acceptable price exists provided that the insurer's expenses are not excessive and that the insurer's risk margin can be sufficiently reduced by diversification within the market segment. This result can be applied to each exposure in the market segment in order to develop supply and demand curves and determine the competitive market price.

Author's note: This paper is based on material presented at the 11th AFIR Colloquium.

References

[1] Borch, K.H. *Economics of Insurance.* Aase, K.K. and Sandmo, A. (Ed.), New York: North-Holland, 1990.

[2] D'Arcy, S.P.; and Dyer, M.A., "Ratemaking: A Financial Economics Approach," PCAS LXXXIV, 1997.

[3] Feldblum, S. "Pricing Insurance Policies: The Internal Rate of Return Model." Exam 9 Study Note, Casualty Actuarial Society. (Available at www.casact.org/library/studynotes/feldblum9.pdf). May 1992.

[4] Grimmett, G.R. and Stirzaker, D.R., *Probability and Random Processes, Second Edition.* Oxford, England: Oxford University Press, 1992.

[5] ISO. "Insurer Financial Results 2001."
http://www.iso.com/studies_analyses/study018.html

[6] Mango, D. "Capital Consumption: An Alternative Methodology for Pricing Reinsurance." Casualty Actuarial Society Forum, Volume: Winter, pp. 351-379, 2003.
http://www.casact.org/pubs/forum/03wforum/03wf351.pdf

[7] Myers, S. and Cohn, R. "A Discounted Cash Flow Approach to Property-Liability Insurance Rate Regulation." *Fair Rate of Return in Property-Liability Insurance*, Cummins, J.D., Harrington, S.A. (Ed.). Norwell, MA: Kluwer Nijhoff Publishing, 1987, 55-78.

[8] Neftci, S.N. An Introduction to the Mathematics of Financial Derivatives. San Diego, California: Academic Press, 1996.

[9] Robison, L.J., and Barry, P.J. *The Competitive Firm's Response to Risk.* New York: Macmillan Publishing Company, 1987.

[10] Schnapp, "The Diversification Property." Presented at the 11th AFIR Colloquium, Toronto, Canada, September 2001.

[11] Schnapp, "Pricing for Systematic Risk." Submitted for the 2004 CAS Ratemaking Conference.

[12] Starmer, C. "Developments in Non-Expected Utility Theory: The Hunt for a Description Theory of Choice under Risk." *Journal of Economic Literature*, XXXVIII, June 2000, 332-382

[13] Thompson, Jr., A. A. and Formby, J. P. *Economics of the Firm, Theory and Practice.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1993.

Appendix A: The Axioms of Expected Utility Theory

Starmer (2000, p. 334) notes that expected utility theory (EUT) can be derived from three axioms: the ordering of preferences, continuity, and independence. Expected utility theory does not require that the outcomes be expressed in monetary terms or even numerically. The only requirement is that the individual must be able to identify a preference between any two uncertain prospects. Taking a more practical viewpoint, it can be assumed that all outcomes are expressed in monetary terms. For the two uncertain random variables X and Y, the notation $X \gg Y$ will represent the individual's assessment that X is a more desirable prospect than Y (i.e., that X is preferred to Y).

The description of preferences and the associated notation will be used here in the normal manner, that is, to represent investments rather than insurance exposures. For insurance exposures, the notation $X \gg Y$ might instead be understood to mean that X is a worse prospect than Y. For instance, if every damage outcome for X exceeds the corresponding damage outcome for Y, then X is a worse prospect than Y. This might be stated as $X \gg Y$ *if* $X(\omega) > Y(\omega)$ for all outcomes ω, provided that all damages are paid at time 0. In order to keep the terminology consistent, insurance damages can be considered to be negative amounts rather than positive amounts. On this basis, the normal terminology for investment decisions would apply as well to insurance transactions.

The ordering axiom requires completeness, i.e., that either $X \gg Y$ or $Y \gg X$, or both. The third possibility is denoted as $X \sim Y$. Ordering also requires transitivity, so that if $X \gg Y$ and $Y \gg Z$, then $X \gg Z$.

The continuity axiom requires that if $X \gg Y$ and $Y \gg Z$, then there exists a probability p such that the compound lottery $(X,p,Z,1-p)$ is equally preferred to Y. The notation $(X,p,Z,1-p)$ represents uncertain damages which are equal to X with probability p or equal to Z with probability $(1-p)$. The two axioms of ordering and continuity imply the existence of a preference function U that assigns a numerical value to each random variable with $U(X) >= U(Y)$ if and only if $X \gg Y$.

The third axiom, that of independence, requires that for all X, Y, and Z, whenever $X \gg Y$ then $(X,p,Z,1-p) \gg (Y,p,Z,1-p)$ for all probabilities p. Starmer (2000, p. 335) observes that "the independence axiom of EUT places quite strong restrictions on the precise form of preferences: it is this axiom which gives the standard theory most of its empirical content (and it is the axiom which most alternatives to EUT will relax)." Further discussion of the independence axiom and the alternatives to expected utility theory can be found in Machina (1987) and Starmer (2000). Given the three axioms, it can be shown that there exists a utility function u such that for all random variables X, the preference function U can be expressed as an expected utility:

$$(21) \qquad U(X) = \sum p(x_i)u(x_i)$$

where $p(x_i)$ represents the probability of outcome x_i.

Appendix B: The Form of the Utility Function

This section will demonstrate that the utility function consisting of two rays, as shown in Figure 3, satisfies equation (12). It will also be shown that the utility function must be of this form.

Assume that the utility function has the form shown in Figure 3. For any damage exposure X, $P(X)$ is the unique value that satisfies the expected utility formula:

$$(22) \qquad EU(w - X + P(X)) = 0$$

Using the definition of U, this can also be expressed as:

$$(23) \qquad -e \int_{x>P(X)} (x - P(X))\, dF(x) + c \int_{x<=P(X)} (P(X) - x)\, dF(x) = 0$$

Consider the random variable $Y = aX$. The expected utility formula for Y is:

$$(24) \qquad EU(w - aX + P(Y)) = -e \int_{ax>P(Y)} (ax - P(Y))\, dF(x) + c \int_{ax<=P(Y)} (P(Y) - ax)\, dF(x)$$

$P(Y)$ is the unique value that makes this result equal to zero. Consider the possibility that $P(Y) = aP(X)$. Substituting this on the right hand side of the formula permits a to be factored out so that:

$$(25) \qquad EU(w - aX + P(Y)) = aEU(w - X + P(X))$$

Since the expected utility formula on the right side of this equation is 0 by definition, this shows that $P(Y) = aP(X)$ is the certainty equivalent price for $Y = aX$, so that $P(aX) = aP(X)$.

The next objective is to demonstrate that U must be of the form shown in Figure 3. Assume that equation (12) is true and consider a random variable X with two outcomes, 1 and x, with $x < 0$. According to the continuity axiom of expected utility theory, there exists a probability p for which $EU(X) = 0$. Hence:

$$(26) \qquad U(x)p + U(1)q = 0$$

so that:

$$(27) \qquad U(x) = -U(1)q/p$$

Next, consider the random variable $Y = 2X$. Since $P(X)$ is the unique solution to $EU(X - P) = 0$, and $EU(X) = 0$, this shows that $P(X) = 0$. Consequently, $P(Y) = 2P(X) = 0$, so that $EU(Y - P(Y)) = EU(Y)$. Since $EU(Y - P(Y)) = 0$ by definition, this means that $EU(Y) = 0$. However:

$$(28) \qquad EU(Y) = U(2x)p + U(2)q = 0$$

or

$$(29) \qquad U(2x) = -U(2)q/p$$

Define $k = U(2) / U(1)$. Since U is concave and increasing, this requires that $1 < k <= 2$. Also, the value for $U(2x)$ is related to the value for $U(x)$ as follows:

(30) $$U(2x) = kU(x)$$

Consider the specific value of $x = -1$, so that $U(-2) = kU(-1)$. Since $U(-1) < 0$ and $k <= 2$, this implies that:

(31) $$U(-2) = kU(-1) >= 2U(-1)$$

However, if $U(-2) > 2U(-1)$, then U is not concave. Consequently, k must be equal to 2, so that for any $x < 0$:

(32) $$U(2x) = 2U(x)$$

A similar result can be obtained for any other random variable $Y = tX$ for $t > 0$, so that:

(33) $$U(tx) = tU(x)$$

The solution to this formula is a straight line ending at the origin:

(34) $$U(x) = ex \quad \text{for } x <= 0$$

A similar argument can be used to show that for $x >= 0$,

(35) $$U(x) = cx \quad \text{for } x >= 0$$

In order for U to be concave, e must be greater than c. U can be converted into the utility function shown in Figure 3 by dividing equations (34) and (35) by c. This shows that the requirement that $P(aX) = aP(X)$ for all $a > 0$ implies that the utility function must be of the form shown in Figure 3.

Appendix C: The Arbitrage Theorem

The arbitrage theorem from the field of financial theory is the basis for developing what is known as arbitrage-free pricing. Arbitrage represents an opportunity to make a risk-free return greater than that of risk-free Treasury bills by taking positions in different assets. If these opportunities do not exist, prices are described as being arbitrage-free. Neftci (1996) describes the arbitrage theorem as providing a connection between risk and the time value of money.

The arbitrage theorem can be described in terms of the following example. Consider three assets, each with a term of one year. The first asset is a bond with a current price of $1 and having a risk-free yield of r_f. The second asset has a current price of S and pays either T - *$1,000* with probability p or T - *$2,000* with probability q, where $p + q = 1$. Both S and T are known values. The third asset is an insurance policy that has a current price of P and pays an indemnity of either $1,000 or $2,000. The Arbitrage Theorem states that an arbitrage-free price P exists for the insurance if and only if there exist positive constants u and v such that:

$$\begin{bmatrix} 1 \\ S \\ P \end{bmatrix} = \begin{bmatrix} 1 + r_f & 1 + r_f \\ T - 1000 & T - 2000 \\ 1000 & 2000 \end{bmatrix} * \begin{bmatrix} u \\ v \end{bmatrix}$$

The matrix equation relates the present value of the three assets to their values one year in the future. The first row represents the ability of the individual to borrow money at the risk-free rate, and can be written as:

$$1 = (1 + r_f) * (u + v) \qquad (36)$$

The values u and v are known as synthetic probabilities. If $y = (1 + r_f) * u$ and $z = (1 + r_f) * v$, it can be seen that y and z are positive values that resemble probabilities. That is:

$$1 = y + z \qquad (37)$$

The second row of the matrix relates the current purchase price for an asset to the value of the uncertain future outcomes. The third row represents the individual's ability to purchase insurance at a premium P to offset the uncertainty of the outcomes for the second asset. The uncertainty can be eliminated since the outcomes for the third asset are negatively correlated with those of the second asset.

Consider the equation describing the second asset:

$$S = (T - 1000) * u + (T - 2000) * v \qquad (38)$$

Multiplying by $(1 + r_f)$, this becomes:

$$S * (1 + r_f) = (T - 1000) * y + (T - 2000) * z \qquad (39)$$

The right hand side of the equation resembles an expected value calculation with the synthetic probabilities y and z used in place of the true probabilities p and q. The transformation of the true probability distribution into a synthetic probability distribution can be understood as a consequence of the risk aversion of the individual.

The following discussion will provide the construction of the synthetic probabilities corresponding to the risk pricing model in equation (4). The initial step is to discount the uncertain outcomes to their present values X using the risk-free rate. The certainty equivalent price P for the exposure X can be expressed as:

(40) $$EU(P - X) = 0$$

or:

(41) $$\int U(P - x)f(x)dx = 0$$

where $f(x)$ is the probability density function of X.

The function $U(P - x)$ is:

(42) $$\begin{array}{ll} (P - x) & for\ P - x \geq 0 \\ (1 + \alpha)(P - x) & for\ P - x < 0 \end{array}$$

Define $g(P) = f(P)$. For all other values, define g(x) as:

(43) $$g(x) = U(P - x)f(x) / (P - x)$$

Since $g(x)$ is non-negative and $\int g(x)dx$ is finite, a synthetic probability density function $h(x)$ can be defined as:

(44) $$h(x) = g(x) / \int g(x)dx$$

Based on these definitions, the expected utility can be restated as:

(45) $$\int U(P - x)f(x)dx = \int (P - x)g(x)dx$$

Since P is the certainty equivalent price, the left side of this equation is 0 so that:

(46) $$\int xg(x)dx = P \int g(x)dx$$

Dividing both sides by $\int g(x)dx$ gives:

(47) $$\int xh(x)dx = P$$

or:

(48) $$E^*(X) = P$$

This result shows that the certainty equivalent price P can be determined as the expected value of the outcomes using the synthetic probability distribution h.

Arbitrage Free Risk Loads For Brokers

Christopher M. Steinbach, FCAS, MAAA

ARBITRAGE FREE RISK LOADS FOR BROKERS

CHRISTOPHER M. STEINBACH

Abstract

When actuaries write about an insurance pricing formula, they first identify the formula in question and then discuss the pros and cons of it. This is in stark contrast to capital markets pricing papers, which start with a list of desirable axioms and then proceed by deriving the necessary and sufficient formula that satisfies these axioms. This paper follows the capital markets paradigm, deriving the necessary and sufficient risk load formula from a set of axioms that are uniquely appropriate for insurance. The derivation follows the same basic approach as Black-Scholes, but it differs in that the axioms have been re-selected to be descriptive of how an efficient insurance market operates. The result is a risk load formula commonly known as the Proportional Hazard Transform. This paper also provides an outline of how to implement the Proportional Hazard Transform, making it relevant and accessible to pricing actuaries.

1. MEASURING RISK

In capital market derivations, 'arbitrage free' means that we can not construct a hedging strategy that is guaranteed to make a profit. This assumption is the key to the dynamic hedging strategies that form the basis for many capital market derivations. The simplest version is the Black-Scholes derivative pricing formula. Variants of Black-Scholes have been derived to overcome the overly simplistic assumptions that Black-Scholes relies upon, but they all follow the same basic construction: they all rely on a dynamic hedging strategy to cancel out the risk.

When it comes to insurance, the main problem with the capital market's arbritrage free approach is that it assumes that the financial institution is not the ultimate risk bearer. This is an important observation because the secret to the Black-Scholes derivation is that it is carefully crafted to make the calculation of risk loads somebody else's problem. Since the dynamic hedging strategy transfers all of the risk to commodity speculators, whether these speculators are adequately compensated for this risk becomes their problem. The existence of speculators who are eager to do the hard work for us is all the derivation requires.

Insurance does not have a separate group of speculators; Insurers and reinsurers fill this role themselves. As a result, insurance pricing formulae must explicitly contain risk loads. So, we can not derive these formulae by pretending they are someone else's problem. Nor can we skirt the issue by creating a virtual commodity market, because we would still need formulae that describe the behavior of the virtual commodity speculators. If we want to derive an insurance pricing formula with an adequate risk load in it, we are going to have to take a different approach.

If we want to borrow the capital market's paradigm, then we need to redefine the term 'arbitrage free' so that it is appropriate for insurance. A more appropriate insurance definition would be that every way we carve up a risk produces the same total risk premium. That is, every two insurance programs that offer the exact same coverage will have the exact same price. In this paper, this insurance version of the arbitrage neutral assumption will be represented by the following axioms:

a. Monotonicity: If $F_X(t) \geq F_Y(t)$ for all t, then the price of insuring X is less than or equal to the price of insuring Y.

b. Scale Invariance: If only p% of a risk is proportionately insured, then the proper premium is p% of the full premium.

c. Layer Invariance: If a risk is carved up into a series of layers, then the sum of the premiums for all of the layers equals the full premium.

d. Continuity: Let X_i be a series of risks that converge to X_∞. [That is, the series of CDFs $F_i(x)$ converge to $F_\infty(x)$.] Then the price of X_i converges to the price of X_∞.

The above axioms are useful to those who care about quantifying what price the market will support, such as insurance brokers and regulators. But, they are not appropriate for risk bearers. Nor are they appropriate when the market is not efficient, such as the market for natural catastrophe reinsurance. In these situations, capital constraints dominate and many of the axioms cease being true.

We can now derive the unique pricing formula that satisfies these axioms. The formula is known as the Proportional Hazard Transform (PHT) and it calculates the risk premium for a risk X. This is the amount of money we need to collect for the expected loss plus the profit load.

PHT Theorem: Let S be a collection of all risks with policy limits of L or less, and let S contain all Bernoulli random variables with payoff of L or less. Then the market premium functional H satisfies Axioms (a)-(d) if and only if H can be represented as:

$$H[X;r] = \int_0^L G^r(t)dt$$

where r has an arbitrary value and G(t) is the decumulative distribution of any risk X in S.

Proof: Appendix A for proof.

Therefore, the risk premium for a risk X is equal to H[X;r], where the parameter r is the risk tolerance parameter. The lower r is, the higher the premium is. When r=1, the premium equals the expected loss. When r=0, the premium equals the policy limit. This gives us the following corollary.

PHT Corollary: r ε [0,1] if and only if H[X;r] ε [E(Loss),Limit] where Limit<∞

Proof: Appendix A for proof.

Note that the restriction that all risks must have a policy limit of L or less can be significantly relaxed, and in many instances removed completely. The restriction is only required for proving the axiom of Continuity from the PHT formula – a relationship with no applied value. Weaker assumptions are possible, but they result in significantly more complicated proofs. These enhancements are left for curious readers to pursue.

It must be reiterated that the PHT is not appropriate for risk bearers. In addition to the weaknesses mentioned above, it should be mentioned that the PHT is not a robust methodology for allocating capital. While an implied capital allocation can be derived from the PHT, it is impossible to create a general formula equating the two that permits the underlying parameters to vary. See Appendix B for an example.

Finally, the above axioms are similar to the axioms originally proposed by Wang, Young & Panjer. The key difference is that the axiom of Comonotonic Additivity has been replaced with Scale and Layer Invariance. The change permits actuaries to better assess the applicability of the axioms as well as permitting a more intuitive proof of the PHT Theorem. See Appendix C for the Wang, Young & Panjer axioms.

2. MEASURING DIVERSIFICATION

While the PHT results in an arbitrage neutral risk premium for any one risk, it does not result in an arbitrage neutral risk premium for a diversified pool of risks. Changing the amount of diversification changes the proper value of the risk tolerance parameter r, as the following examples demonstrate.

Example #1: Calculate H[X;r] for X=B(p,1), a Binomial risk of probability p and payout $1.

$$f[X]=\begin{cases}0 & \text{prob } 1-p\\ 1 & \text{prob } p\end{cases}$$

$$F[X]=\begin{cases}1-p & x<1\\ 1 & x\geq 1\end{cases}$$

$$G[X]=\begin{cases}p & x<1\\ 0 & x\geq 1\end{cases}$$

$$H[X;r] = \int_0^L G^r(t)dt = p^r$$

Example #2: Calculate H[X+Y;r] for X=Y=B(p,1), i.i.d.

$$f[X]=\begin{cases} 0 & \text{prob.}\,(1-p)^2 \\ 1 & \text{prob.}\,2p-2p^2 \\ 2 & \text{prob.}\,p^2 \end{cases}$$

$$F[X]=\begin{cases} (1-p)^2 & x<1 \\ 1-p^2 & 1\le x<2 \\ 1 & x\ge 2 \end{cases}$$

$$G[X]=\begin{cases} 2p-p^2 & x<1 \\ p^2 & 1\le x<2 \\ 0 & x\ge 2 \end{cases}$$

$$H[X;r] = \int_0^L G^r(t)dt = (2p-p^2)^r + p^{2r}$$

The above examples demonstrate that it is more expensive to insure two risks separately than it is to insure them together, assuming the same risk tolerance parameter is used for both calculations. This can be shown by comparing the price of insuring two identical Bernoulli risks separately, as in example #1, against insuring the risks together, as in example #2. If the two approaches produced identical risk premiums, then we would have:

$$2p^r = (2p-p^2)^r + p^{2r} \qquad \text{for all p}$$

which we can rewrite as:

$$(2-p^r) = (2-p)^r \qquad \text{for all p}$$

But, this is true only when r equals 0 or 1. That is, this is true only if the premium equals the expected loss or the full limit – the two extremes.

The conclusion is that there are no natural values for the risk tolerance parameter other than the two extremes. Risk tolerance is necessarily a relative measure that reflects the degree of portfolio diversification that a risk benefits from. The better diversified a risk is, the greater the market's risk tolerance for that risk.

For this reason, the PHT does not actually permit actuaries to compute the risk premium for a risk from first principals. Instead, it requires the actuary to first establish the correct risk tolerance parameter for that risk by determining what the market supports, and then it calculates the risk premium from that value. The result is a price that is consistent with the prevailing state of the market. This is the price that the market is expected to be able to bear.

There are several ways to calibrate the risk tolerance parameter r. The simplest approach is to keep a table of risk tolerance values for comparable risks. These are what the market has supported for similar risks. For example:

Example #3: Suppose we have just bound a quota share treaty where the market supported a 60% loss ratio pick and a 65% risk premium. The empirically derived loss distribution is listed below. Then the corresponding r is determined by finding the value that results in the risk premium. The results are as follows:

X	f(X)	F(X)	G(X)	H(X)
0%	0.0%	0.0%	100.0%	100.0%
10%	0.0%	0.0%	100.0%	100.0%
20%	0.0%	0.0%	100.0%	100.0%
30%	0.0%	0.0%	100.0%	100.0%
40%	10.0%	10.0%	90.0%	94.0%
50%	20.0%	30.0%	70.0%	81.1%
60%	40.0%	70.0%	30.0%	49.2%
70%	20.0%	90.0%	10.0%	25.8%
80%	10.0%	100.0%	0.0%	0.0%
90%	0.0%	100.0%	0.0%	0.0%
Sum	100.0%		600.0%	650.0%

Now, since the width of each layer is 10%, the expected loss is 10% of the sum of G(X) and the risk premium is 10% of the sum of H(X). Note that $H(X)=G^r(X)$ in this table. The value that resulted in the risk premium equaling 65% was r=0.589. This value then gets put into a table that lists the name of the account, a description of the business, the date of inception and the value of r that the market supported.

Note that when constructing tables of r's, different reinsurance products for the same line of business can have different r's. That is, r is not always the same from product to product even when the underlying business is the same. There are several reasons for this:

1. Different reinsurance products are supported by different sectors of the reinsurance market. Each sector has a different amount of capacity available to it, resulting in differences in supply by product.
2. Different reinsurance products have different regulatory and statutory consequences. Some are treated more favorably than others, resulting in differences in demand by product.
3. Large and small losses can come from different causes of loss, each with a different parameter variance. Thus, parameter variance can vary by product.

Finally, when a risk is carved up, we are faced with the problem that the sum of the parts may not equal the whole. Equality holds for programs carved up vertically and horizontally, but it does not hold for other structures. Notably, when we refer to "Layer Invariance," we are referring to aggregate stop loss layers only. It does not hold for excess reinsurance because Layer Invariance slices up the loss distribution after frequency has been reflected. The result is that excess reinsurance structured as two narrow layers costs slightly more than one wide layer, as the following example demonstrates.

Example #4: Suppose we are analyzing a line of business which has a frequency that is Poisson with a mean of 10 and a severity that is Exponential with a mean of $1M. We are interested in pricing the $1M x $1M, $3M x $2M and $4M x $1M excess layers. If r=0.589 then the risk premium for the $1M x $1M, $3M x $2M and $4M x $1M layers are $2,842,000, $2,049,000 and $4,740,000 respectively. In other words, reinsuring both layers together is 3.1% cheaper than reinsuring them separately. (See Appendix D for the calculation details.)

There are times when this inequality should be compensated for, such as when reinsurers all take the same shares on two consecutive excess layers. For example, if the reinsurers in Example #4 behave this way, then the pricing of the individual layers should be multiplied by a compensation factor of 0.969 = $4,740,000 / ($2,842,000 + $2,049,000) to force the sum of the parts to equal the whole.

When pricing any structure, the actuary should always check the final results to see if diversification is being properly applied. In many situations, diversification is not being properly applied and a compensation factor is required. Also, when borrowing r's from comparable reinsurance programs, be careful what you consider a comparable program to be. Not only must the underlying risk be comparable, the market's acceptance of the product must also be comparable.

3. MEASURING DURATION

A practical risk load formula has three components to it: a measure of risk, a measure of diversification, and a measure of duration. The PHT formula includes a measure of risk and the compensation factor corrects for diversification, but we have yet to reflect duration. The main complication we must address is that risk can run off very differently from expected loss. For example, the fact that 50% of the loss has incurred does not mean that 50% of the risk has occurred.

We can model the run-off of risk by borrowing ideas from IBNR theory. In IBNR theory, the Paid Loss Lag at time t is defined to be the percentage of the total expected loss that we expect to have been paid by time t. Similarly, the Incurred Loss Lag at time t is defined to be the percentage of the total expected loss that that we expect to have incurred by time t. We can define the Risk Lag at time t as the percentage of the total expected risk that we expect to have occurred by time t.

The concept of a Risk Lag permits us to use the Paid Lags and Incurred Lags as benchmarks for estimating Risk Lags. First notice that no risk remains after a loss has been fully paid. Therefore, the Paid Lags can serve as an approximate upper bound for the Risk Lags. Also notice that all of the risk remains until the claim has been first reported. Therefore, the Incurred Lags can serve as an approximate lower bound on the

Risk Lags. This permits us to estimate Risk Lags as weighted averages of Paid and Incurred Lags.

Where exactly the Risk Lags lie between the Paid Lags and Incurred Lags would depend on how claims are handled. If case reserves are always established with 100% accuracy, then the only risk occurs when the loss occurs, so the Risk Lag should equal the Incurred Lag. If case reserves have zero credibility until the loss is paid, then risk does not diminish until the loss is paid and the Risk Lag should equal the Paid Lag. An intermediate reserving approach would result in an intermediate Risk Lag.

While this model of duration is not exceptionally precise, it does produce reasonable and practical estimates. Its strength is that it permits actuaries to make intelligent estimates of how risk runs-off and it permits relatively sophisticated calculations of the cost of capital. More complicated structures can be created to model the time value of risk more accurately, but these are left for ambitious readers to pursue.

4. CONCLUSION

The PHT calculates risk premiums that are consistent with what an insurance market bears, reflecting the market's tolerance for risk. This makes the PHT especially useful to brokers who are interested in predicting the price that the market should support. But, the PHT assumes that an efficient insurance market exists. To the extent that an efficient market does not exist, the result of the PHT will not be accurate. None-the-less, the PHT derived price should be a lower bound on the market price, permitting benchmarking even when the market is not efficient.

The PHT is significantly less useful to risk bearers. This can be seen in the way the axioms do not reflect the risk bearer's capital structure. Risk bearers can use the PHT formula to estimate what the market should bear for a risk, but they should not use the formula to establish if this risk is an effective use of their capital. Other formulae, such as Value at Risk, are more appropriate for this purpose.

Interestingly, while the PHT does not reflect the capital structure of an individual company, it does make assumptions about the capital utilization of the industry as a whole. This is evident in the fact that the resulting risk loads are not zero, indicating that the industry does have capital at risk. But, instead of modeling capital as a fixed quantity the way individual insurers do, the PHT assumes that market forces will require insurers to actively manage their capital to match the risk they assume. In other words, the companies who use their capital efficiently will determine the market and everyone else must struggle to keep up.

Considerably more work needs to be done on pricing & capital allocation. We have historically assumed that the Law of Large Numbers makes our industry different and our pricing formulae unique. But, it is increasingly obvious that insurance is a form of finance and that the financial markets paradigms should be applied to insurance. Hopefully, this paper contributes to the change in the direction of actuarial discourse, closing some of the divides that currently exist between insurance and the capital markets.

REFERENCES

[1] Feldblum, Sholom, "Risk Loads for Insurers," PCAS LXXVII, pp.160-195

[2] Rudin, Walter, "Real & Complex Analysis," McGraw-Hill Book Company, 1987

[3] Wang, Shaun, "Insurance Pricing and Increased Limits Ratemaking by Proportional Hazard Transforms," Insurance: Mathematics and Economics 17, 1995, pp. 43-54

[4] Wang, Shaun, "Premium Pricing by Transforming the Layer Premium Density," ASTIN 26, 1996, pp. 71-92

[5] Wang, Shaun, Virginia Young and Harry H. Panjer, "Axiomatic Characterization of Insurance Price," Insurance: Mathematics and Economics 21, 1997, pp.173-183

[6] Wheeden, Richard L., and Zygmund, Antoni, "Measure and Integral, An Introduction to Real Analysis," Marcel Dekker, Inc., 1997

APPENDIX A

PROOF OF THE PHT THEOREM & COROLLARY

Lemma 1: Let B(q,k) represent a Bernoulli risk with probability q and severity k. Let H(q,k) represent the premium we would charge for insuring this risk. The unique formula for H(q,k) that satisfies axioms (a), (b) & (e) is the following:

$$H(q,k) = q^r k \qquad [1]$$

where r is an arbitrary risk tolerance parameter.

Proof: Let us model B(q,k) as a biased coin toss, where the coin has a probability p of heads (p is not necessarily equal to q in [1]). Consider the following process: Flip a biased coin. If heads, flip it again. If heads again, then collect $k. This process can be insured in two ways. First, it can be considered a Bernoulli risk with p^2 probability of heads and a pay-out of k. Second, it can be considered as a Bernoulli risk with probability p and a pay-out that buys another insurance policy insuring against the second coin toss. Since these two insurance structures provide the exact same protection, the axiom of Monotonicity says that they must have the exact same price. Thus, we can equate the risk premiums for these two different insurance structures:

$$H(p^2,k) = H(p,H(p,k)) \qquad [2]$$

Furthermore, the axiom of Scale Invariance implies that the premiums are linearly proportional to the payoff:

$$H(p,X) = X\,H(p,1) \qquad [3]$$

Applying [3] to [2] we get:

$$H(p^2,1) = H^2(p,1) \qquad [4]$$

For the sake of notational simplicity, the k's have been cancelled out. We will reintroduce them at the end. Applying the above argument repeatedly gives us:

$$H(p^m,1) = H^m(p,1) \text{ for all positive integers m} \qquad [5]$$

Which we can rewrite as:

$$H^{1/m}(p^m,1) = H(p,1) \text{ for all positive integers m} \qquad [6]$$

Letting n be another integer (not necessarily equal to m), we can rewrite [6] with n replacing m. Equating the two versions of [6] gives us:

$$H^{1/m}(p^m,1) = H^{1/n}(p^n,1) \text{ for all positive integers m,n} \qquad [7]$$

or

$$H(p^m,1) = H^{m/n}(p^n,1) \text{ for all positive integers m,n} \qquad [8]$$

Now, let $q=p^n$ and substitute this into [8]:

$$H(q^{m/n},1) = H^{m/n}(q,1) \text{ for all positive integers } m,n \quad [9]$$

Applying the Continuity assumption gives us:

$$H(q^v,1) = H^v(q,1) \text{ for all positive real numbers } v \quad [10]$$

Which we can write as:

$$H^{1/v}(q^v,1) = H(q,1) \text{ for all positive real numbers } v \quad [11]$$

Now, since v can be anything, let us select v such that $q^v = 0.5$. In other words, $v=\ln(0.5)/\ln(q)$. Substituting this into [11]:

$$H(q,1) = H^{\ln(q)/\ln(0.5)}(0.5,1) \quad [12]$$

Which we can rewrite as:

$$H(q,1) = q^{\ln(H(0.5,1))/\ln(0.5)} \quad [13]$$

Letting $r = \ln(H(0.5,1)) / \ln(0.5)$, reintroducing k and applying [3] again, we get:

$$H(q,k) = q^r k \quad [14]$$

which is the PH-Transform for B(q,k).

Lemma 2: Let H[Z] represent the premium charged for an arbitrary risk Z. Let G(z) be the decumulative distribution of Z. The unique H[Z] that satisfies axioms (a) - (e) is the following:

$$H[Z] = \int_0^\infty G^r(t)\,dt \quad [15]$$

Proof: By the Layer Invariance assumption, the price for risk Z is the sum of the prices for each of its layers. Carve the risk Z into a series of layers, each of width i. Let j enumerate the layers with j=0 being the first layer. Now we must evaluate the prices of the layers.

First, by the axiom of Monotonicity, the price of excess layer number j is bounded from above by a Bernoulli risk of probability G(ji) and severity of i. [Probability of entering the layer & severity of one layer width. This prices every partial loss in the layer as if it were a total loss, resulting in a price that is too high.] So we can write:

$$H[Z] \le \sum_{j=0}^{\infty} H[G(ji), i] \quad [16]$$

The Continuity assumption permits us to take the limit as i approaches zero giving us:

$$H[Z] \le \lim_{i\to 0} \sum_{j=0}^{\infty} H[G(ji), i] \qquad [17]$$

Applying Lemma 1 and evaluating:

$$H[Z] \le \int_0^{\infty} G^r(t)\,dt \qquad [18]$$

Second, by the axiom of Monotonicity, the price of excess layer j is bounded from below by a Bernoulli risk of probability G((j+1)i) and severity i. [Probability of exhausting the layer & severity of one layer width. This ignores partial losses entirely, resulting in a price that is too low.] So, we can write:

$$H[Z] \ge \sum_{j=0}^{\infty} H[G((j+1)i), i] \qquad [19]$$

The Continuity assumption permits us to take the limit as i approaches zero giving us:

$$H[Z] \ge \lim_{i\to 0} \sum_{j=0}^{\infty} H[G((j+1)i), i] \qquad [20]$$

Applying Lemma 1 and evaluating:

$$H[Z] \ge \int_0^{\infty} G^r(t)\,dt \qquad [21]$$

Equations [18] and [21] together produce [15].

Lemma 3: Let $\{G_k^r\}$ be a sequence of measurable functions on E=[0,L] such that $G_k^r \to G^r$ a.e. in E. Since $|E|<+\infty$ and $|G_k^r| \le 1$ a.e. in E, then $\int_E G_k^r \to \int_E G^r$.

Proof: This is a specific case of the Bounded Convergence Theorem. See any Real Analysis textbook (such as Wheeden & Zygmund) for the proof.

PHT Theorem: Let S be a collection of all risks with policy limits of L or less and let S contain all Bernoulli random variables with payoff of L or less. Then the market premium functional H satisfies Axioms (a) - (e) if and only if H can be represented as

$$H[X;r] = \int_0^L G^r(t)dt \qquad [18]$$

where r has an arbitrary value and G(t) is the decumulative distribution of any risk X in S.

Proof:

$\Rightarrow$ Apply Lemma 2

$\Leftarrow$ The PHT has the following properties:

a) If $F_X(t) \geq F_Y(t)$ for all t, then $G_X(t) \leq G_Y(t)$ for all t and the result immediately follows.

b) The price of a risk X is:

$$H[X;r] = \int_0^\infty G^r(t)dt \qquad [19]$$

Let $v=p\cdot t$ and $dv=p\cdot dt$, then

$$p \cdot H[X;r] = \int_0^\infty G^r\left(\frac{v}{p}\right)dv \qquad [20]$$

And since the integral is H[pX;r], this proves Scale Invariance.

c) The price for the layer that starts with A and ends with B equals:

$$H[X;r] = \int_A^B G^r(t)dt \qquad [21]$$

Layer Invariance immediately follows.

d) Apply Lemma 3.

PHT Corollary: $r \in [0,1]$ if and only if $H \in [E(\text{Loss}),\text{Limit}]$ where $\text{Limit}<\infty$

Proof:

$\Rightarrow$ If r=0, then H=Limit. If r=1, then H=E(loss). All that is left is to prove H is strictly monotonic. We know that $0 \leq G(t) \leq 1$ for all t. Therefore, if r<s, then $G^r(t)>G^s(t)$ and $H[X;r]>H[X;s]$ for all $H\neq\infty$. In our situation, H is bounded above by Limit, so H is necessarily finite.

$\Leftarrow$ Since all finite H are strictly monotonic with respect to r, there is only one r that results in H=Limit and only one r that results in H=E(Loss). We know from above that these are r=0 and r=1 respectively. Application of strict monotonicity completes the proof.

APPENDIX B

IMPLIED CAPITAL ALLOCATION

While the PHT can be used for calculating an implied capital allocation, it is not a robust approach. For example, a Bernoulli risk with a probability of p and a payout of $1 has an expected loss of p and a risk premium of p^r. This means that the load for risk is p^{r-1}. If the market supports an ROE of R, then the implied capital allocation is p^{r-1}/R.

The problem occurs when the insurance market's pricing changes (either getting harder or softer). Continuing the above example, if we assume that the ROE changes to R' while the capital allocation stays the same, then the new price H' for the risk is:

$$H' = p + p^{r-1} \times R' / R = p^{r'}.$$

Notice that the capital allocation is a function of r & R and the new risk tolerance parameter r' is a function of p. This is the problem. If the PHT is equivalent to a capital allocation methodology, then we should have a capital allocation that is a function of only p and the cost of capital should be invariant with respect to p. Instead, the PHT results in the two being commingled. It appears to be impossible to disaggregate the PHT's risk load into an amount of capital formula times a cost of capital formula.

For example, suppose we have three $1 Bernoulli risks with p equal to 25%, 50% and 75% respectively. If r=0.85, then the risk pure premiums for the three risks are 0.308, 0.555 and 0.783 respectively. If R=10%, then the implied capital allocations are 0.578, 0.548 and 0.331 respectively. If the ROE now changes to 15%, then the new risk pure premiums are now 0.337, 0.582 and 0.800 respectively. But, this results in the new risk tolerance parameter varying by risk. The risk tolerance parameters for the three risks are now 0.785, 0.780 and 0.777 respectively. Notice that the only way to make the new risk tolerance parameter constant is to either vary the new ROE or to change the capital allocations. The risk tolerance parameter, the ROE and the capital allocation can not all be invariant.

Therefore, the PHT approach to pricing and the capital allocation approach to pricing are incompatible. As the insurance market gets harder or softer, one of the following must be true:

1. The PHT's risk tolerance parameter r varies by risk
2. The market's ROE varies by risk
3. The capital allocations change in addition to the cost of capital changing

In other words, the assumptions underlying the two pricing methodologies can not all be true. Further research must be performed to determine which set of assumptions better explains the behavior of the insurance market.

APPENDIX C

THE WANG-YOUNG-PANJER AXIOMS

In 1997, Wang, Young and Panjer published the paper "Axiomatic Characterization of Insurance Price" in which they identified and proved a set of necessary and sufficient conditions for the PHT. The axioms Wang, Young and Panjer identified were as follows:

1. Conditional State Independence: For a given market condition, the price of an insurance risk X depends only on its distribution.
2. Monotonicity: For two risks X and Y in $\mathcal{X}$, if $X(\omega) \leq Y(\omega)$, for all $\omega \in \Omega$, a.s., then $H[X] \leq H[Y]$.
3. Comonotonic Additive: if X and Y in $\mathcal{X}$ are comonotonic, then $H[X+Y] = H[X]+H[Y]$.
4. Continuity: For $X \in \mathcal{X}$ and $d \geq 0$, the functional H satisfies

   $$\lim_{d \to 0+} H[(X - d)_+] = H[X] \quad \text{and} \quad \lim_{d \to \infty} H[\min(X, d)] = H[X]$$

   in which $(X - d)_+ = \max(X - d, 0)$.
5. Reduction of compound Bernoulli risks: Let $X = IY$ be a compound Bernoulli risk, where the Bernoulli frequency random variable I is independent of the loss severity random variable $Y = X \mid X>0$. Then the market prices for risks $X = IY$ and $IH[Y]$ are equal.

From these axioms, Wang, Young and Panjer prove the following:

W.Y.P's PHT Theorem: Assume that a collection of risks $\mathcal{X}$ contains all of the Bernoulli random variables. Then the market premium functional H satisfies Axioms 1-5 if and only if H can be represented as

$$H[X;r] = \int_0^\infty G^r(t)dt$$

where r is some unique positive constant and G(t) is the decumulative distribution of any risk in $\mathcal{X}$.

APPENDIX D

EXCESS REINSURANCE EXAMPLE

Assume that frequency is Poisson with a mean of 10, severity is Exponential with a mean of $1M, and r=0.589. We want to price the excess reinsurance layers $1M x $1M, $3M x $2M and $4M x $1M. First we determine the decumulative distribution G(X) by simulating 10,000 years of experience. Then we calculate H(X):

$1M x $1M Excess Layer

Discretization Point X	Midpoint	G(X)	H(X)
-	500,000	89.1%	93.4%
1,000,000	1,500,000	62.0%	75.5%
2,000,000	2,500,000	33.3%	52.3%
3,000,000	3,500,000	14.4%	31.9%
4,000,000	4,500,000	4.9%	16.9%
5,000,000	5,500,000	1.3%	7.7%
6,000,000	6,500,000	0.4%	3.9%
7,000,000	7,500,000	0.2%	2.6%
8,000,000	8,500,000	0.0%	0.0%
Sum		205.6%	284.2%
Layer Width		1,000,000	1,000,000
Expected Loss		2,056,000	
Risk Premium			2,842,000

$3M x $2M Excess Layer

Discretization Point X	Midpoint	G(X)	H(X)
-	500,000	53.8%	69.4%
1,000,000	1,500,000	30.4%	49.6%
2,000,000	2,500,000	16.6%	34.7%
3,000,000	3,500,000	6.9%	20.7%
4,000,000	4,500,000	3.2%	13.2%
5,000,000	5,500,000	1.4%	8.1%
6,000,000	6,500,000	0.6%	4.9%
7,000,000	7,500,000	0.2%	2.6%
8,000,000	8,500,000	0.1%	1.7%
Sum		113.2%	204.9%
Layer Width		1,000,000	1,000,000
Expected Loss		1,132,000	
Risk Premium			2,049,000

$4M x $1M Excess Layer

Discretization Point X	Midpoint	G(X)	H(X)
	500,000	89.1%	93.4%
1,000,000	1,500,000	71.7%	82.2%
2,000,000	2,500,000	54.1%	69.6%
3,000,000	3,500,000	38.7%	57.2%
4,000,000	4,500,000	26.2%	45.4%
5,000,000	5,500,000	16.9%	35.1%
6,000,000	6,500,000	10.1%	25.9%
7,000,000	7,500,000	6.2%	19.4%
8,000,000	8,500,000	3.4%	13.6%
9,000,000	9,500,000	1.9%	9.7%
10,000,000	10,500,000	1.0%	6.6%
11,000,000	11,500,000	0.6%	4.9%
12,000,000	12,500,000	0.3%	3.3%
13,000,000	13,500,000	0.2%	2.6%
14,000,000	14,500,000	0.1%	1.7%
15,000,000	15,500,000	0.1%	1.7%
16,000,000	16,500,000	0.1%	1.7%
17,000,000	17,500,000	0.0%	0.0%
Sum		320.7%	474.0%
Layer Width		1,000,000	1,000,000
Expected Loss		3,207,000	
Risk Premium			4,740,000

Note that, in the above calculations, the discretization includes a point at the origin. The probability assigned to a discretization point equals the cumulative probability from the midpoint directly below the discretization point to the midpoint directly above the descretization point. For example, the probability assigned to the $2.0M discretization point is the cumulative probability that the loss will be between $1.5M and $2.5M. This discretization process results in an expected loss that is slightly lower than the simulated expected loss. For example, the $4M x $1M layer expected loss is calculated above to be $3,207,000 while it the simulation really resulted in an expected loss of $3,211,307.

Also note that the values of G(X) and H(X) are both rounded to the first decimal point, as shown above. This was done solely to make the calculations easier for readers to follow.

Credible Risk Classification—How To Create Risk Classification Systems With the Maximum Price Differentiation While Addressing Concerns of Credibility

Benjamin Joel Turner, MBA, JD

Credible Risk Classification

How to create risk classification systems with the maximum price differentiation while addressing concerns of credibility.

By

Benjamin Joel Turner[1] MBA, JD

Abstract

"Ratemaking accuracy often is improved by subdividing experience into groups exhibiting similar characteristics... There is a point at which partitioning divides data into groups too small to provide credible patterns." (*Statement of Principles Regarding Property and Casualty Insurance Ratemaking*, adopted by the Board of Directors of the CAS, May 1988, page 3.)

This paper outlines an objective approach to creating classification systems with maximum price differentiation while addressing concerns of credibility. The formula developed in this paper will yield a numeric score, which ranks the varying classification systems as applied to a given data set. The score is capable of ranking class plans with different numbers of classes. Therefore, this formula determines "the point at which partitioning divides data into groups too small" to be of value for price differentiation in a competitive marketplace.

[1] Benjamin_Turner@hotmail.com, (In order to avoid my filter, the subject must contain the word "actuary")

1. MOTIVATION

I recently was employed on a team that evaluated and offered suggestions concerning the automobile underwriting guidelines for one of the major US automobile insurers. During this process we invented new class plans and compared them to each other and to the current plan. When alternatives had the same number of classes, our team was able to determine which plan was superior; we merely determined which plan had the highest variance of the means, as detailed in Woll [1] and also in Finger [2].

When alternative class plans were compared with different numbers of classes, we were unable to agree upon an objective method for comparing the different plans. When an alternative contained "too many" groupings, or one of the groups had "few" exposures, we feared that we were losing credibility. Another approach was to test for "significance," but we were unable to agree upon a significance level.

The goal of our team was to provide the maximum price differentiation. Companies that can successfully do so will be able to skim the cream from the total pool of potential insureds. To provide maximum price differentiation, there is a temptation to continually subdivide the book of business into an ever-increasing number of classes. However, common sense dictates that there is a point where the book is so segmented that it is no longer credible. When the book is over-segmented, its predictive power is so poor that instead of skimming the cream, the insurance company is haphazardly selecting insureds for discounts that are not truly warranted. Furthermore, a company may choose to credibility-weigh each class against the overall book mean. Due to the fact that an over-segmented book has low credibility, the complement of credibility overpowers the initial class mean, and the goal of price segmentation is thwarted. This paper is an attempt to find the point at which further segmentation is counterproductive.

2. THE METHOD

An objective way to balance the need for price differentiation (as measured by the variance of the means) with the need for credibility is to first credibility-weigh the means and then calculate the variance of the credibility-weighted means. The Nonparametric-Buhlman-Empirical-Bayes Method [3] is used to calculate the credibility (see Appendix A for formulas).

This method[2] is particularly advantageous for this purpose because:

- It assumes no underlying distribution.
- It is relatively uncontroversial.
- It supplies its own complement of credibility.
- It does not require arbitrary selection of parameters.

The comparison is accomplished by computing the variance of the credibility-weighted means for each alternative plan and selecting the plan with the highest value. This method allows for direct comparison of simple plans vs. complex plans (for example a simple plan of four classes versus a complex plan with eleven classes). It is not

[2] When credibility is low, this method yields a negative value, which practitioners are advised to set to zero. This aspect of the method is useful because it allows for a balance between credibility and explanatory

necessary to arbitrarily designate one plan as the null hypothesis, select a confidence interval, or assume a statistical distribution.

3. A SIMPLE EXAMPLE

3.1 Introduction

Suppose there is a rating factor with four levels that are increasing in risk. These levels could be classified in the following ways: 1, 2-3-4; 1-2, 3-4; 1-2-3, 4; 1-2, 3, 4; 1, 2-3, 4; 1, 2, 3-4; 1-2-3-4; 1, 2, 3, 4. Which class plan has the most explanatory power? The standard variance of the means technique will always choose the latter class plan (four distinct classes), even if all of the levels have the same underlying mean. This method does not account for credibility, and thus will always favor the maximum number of classes.

Suppose the data in Table 3.1.1 below.

Table 3.2.1

Class	PolCounts[3]	Exposures	Losses	Losses/Exposures	LossesSquared[4]
1	1,140	1,741	1,265,754	727	12,882,705,642
2	1,000	1,514	1,390,038	918	16,157,016,292
3	960	1,456	1,359,435	934	15,132,695,151
4	1,060	1,609	1,846,048	1,147	22,805,214,328

power. When credibility is low it will be set to zero, at which point explanatory power, as measured by the variance of the credibility-weighted means, will also be zero.

[3] This will always denote the count of the policies.

[4] For each policy the losses are squared and then divided by the exposures of that policy. The results are summed up by level.

Should each level ultimately be assigned to its own class, or would it be better to group the levels? For example, Levels 2 and 3 are quite similar in Losses per Exposure. Should they be consolidated into one class?

3.2 Calculation of Credibility

Tables 3.2.1 and 3.2.2 show how to calculate credibility for Table 3.1.1 where each level is assigned as its own class.

Table 3.2.1

Class	Polcounts	Exposures	Losses	LossesSquared[4]	Exposures^2	Losses^2 / Exposures	W	Credibility
1	1,140	1,741	1,265,754	12,882,705,642	3,031,081	920,237,328	69,911,815	**0.721**
2	1,000	1,514	1,390,038	16,157,016,292	2,292,196	1,276,225,655	130,779	**0.692**
3	960	1,456	1,359,435	15,132,695,151	2,119,936	1,269,274,395	57,075	**0.683**
4	1,060	1,609	1,846,048	22,805,214,328	2,588,881	2,118,019,402	77,811,443	**0.705**
Sum		6,320	5,861,275	66,977,631,413	10,032,094		147,911,112	

$$W_i = Exposures_i * \left(\frac{Losses_i}{Exposures_i} - \frac{\sum Losses_i}{\sum Exposures_i}\right)\text{^}2$$

$$Credibility_i = \frac{Exposures_i}{Exposures_i + K}$$

"K" is calculated below.

Table 3.2.2

Other Calculated Values	
V	14,772,347
A	21,889
K = V/A	675

$$\text{V} = \left\{\frac{1}{\sum(\text{PolicyCounts}_i - 1)}\right\} * \left\{\left(\sum \text{LossesSquared}_i\right) - \left(\sum \frac{\text{Losses}_i \text{^}2}{\text{Exposures}_i}\right)\right\}$$

$$\text{A} = \frac{(\sum \text{W}_i) - \text{V} * (\text{Number of Classes - 1})}{(\sum \text{Exposures}_i) - \frac{\sum \text{Exposures}_i \text{^}2}{\sum \text{Exposures}_i}}$$

Consolidation

In contrast to the illustration above, levels two and three are consolidated into one class so that there are three classes in the plan. Then the data is as in Tables 3.2.3 below.

Table 3.2.3

Class	PolCounts	Exposures	Losses	LossesPerExposure	LossesSquared
1	1,140	1,741	1,265,754	727	12,882,705,642
2-3	1,960	2,970	2,749,473	926	31,289,711,443
4	1,060	1,609	1,846,048	1,147	22,805,214,328

Using the same methodology as above, the calculation of credibility is shown in Table 3.2.4 and Table 3.2.5 below.[5]

Table 3.2.4

Class	Polcounts	Exposures	Losses	LossesSquared	Losses^2 / Exposures	Exposures^2	W	Credibility[6]
1	1,140	1,741	1,265,754	12,882,705,642	920,237,328	3,031,081	69,911,815	**0.775**
2-3	1,960	2,970	2,749,473	31,289,711,443	2,545,320,464	8,820,900	8,268	**0.855**
4	1,060	1,609	1,846,048	22,805,214,328	2,118,019,402	2,588,881	77,811,443	**0.761**
Sum		6,320	5,861,275	66,977,631,413		14,440,862	147,731,526	

Table 3.2.5

Other Calculated Values	
V	14,768,837
A	29,292
K	504

[5] Note that a separate data run is not necessary. If the data contains the field called "LossesSquared," the levels can be consolidated or divided at will, and credibility can be recalculated.

[6] Note that Credibility for Class 1 and Class 4 is higher in the consolidated plan even though the number of exposures for these Classes is the same.

3.3 Calculation of Score

Non-consolidated Class Plan

Once credibility has been assigned, Score is calculated thus, where "i" is the number of classes:

$$\text{Score} = \frac{\text{Explained Credible Variance}}{\text{Total Variance}}$$

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{BookMean} - \text{Cred.Wtd.ClassMean}_i)^\wedge 2}{(\sum_i \text{LossesSquared}_i) - (\text{Exposures} \times [\text{BookMean}]^\wedge 2)}$$

Table 3.3.1 shows the calculation of the numerator in Score.

Table 3.3.1

Class	Book Mean	Avg. Loss per Class	Credibility-weighted Class Mean	Difference	Squared	Multiplied by Exposures
1	927	727	783	(144)	20,855	36,307,970
2	927	918	921	(6)	41	62,568
3	927	934	932	4	18	26,647
4	927	1,147	1,082	155	24,003	38,620,138
Sum						**75,017,324**

The denominator is 66,977,631,413 – 6,320 * (927^2) = **61,546,672,133**.

Score is 75,017,324 / 61,541,785,744 = **0.122%**[7]

[7] Some may be concerned that Score is such a low percentage, but this evinces the importance of insurance. If a class plan could yield a Score of 100%, then the future would be entirely predictable, and insurance would be unnecessary.

Consolidated Class Plan

Table 3.3.2 shows the calculation of the numerator for the consolidated class plan.

Table 3.3.2

Class	Book Mean	Avg. Loss per Class	Credibility-weighted Class Mean	Difference	Squared	Multiplied by Exposures
1	927	727	772	(155)	24,146	42,037,707
2-3	927	926	926	(1)	2	6,042
4	927	1,147	1,095	167	28,036	45,110,265
Sum						**87,154,014**

The denominator is the same as in the prior example, **61,546,672,133**.

Score is 87,154,014/ 61,541,785,744 = **0.142%.**

(See Appendix A for a recap of the formulas.)

All Possible Class Plans

Similar calculations were performed for all of the possible class plans and are shown in Table 3.3.3:

Table 3.3.3

Class	Score
1-2-3-4 (no segmentation)	0.000%
1-1, 2-4	0.107%
1-2, 3-4	0.092%
1-3, 4-4	0.118%
1-1, 2-2, 3-4	0.104%
1-1, 2-3, 4-4	0.142%
1-2, 3-3, 4-4	0.110%
1, 2, 3, 4	0.122%

3.4 Interpretation of Score

The above values of Score imply that after considering credibility, the class plan called "1, 2-3, 4" has a greater ability to provide price differentiation than the class plan called

"1, 2, 3, 4," or any other class plan. Thus, in this example, three classes produce more price differentiation than four classes after adjusting for credibility.

3.5 Comparison to Hypothesis Testing

Hypothesis testing can also be used to determine if the simple example should have been grouped as 1, 2, 3, 4; 1, 2-3, 4; or any other class plan. Using the rationale identified by Salam [4], levels are divided only if they are statistically significant. For this example,[8] the standard test for difference of means is used to determine statistical significance:

$$Z = \frac{(ClassMean_A - ClassMean_B)}{\sqrt{\frac{Variance_A}{Exposures_A} + \frac{Variance_B}{Exposures_B}}} = \frac{\text{Difference of Means}}{\text{Standard Deviation}} \sim \text{Standard Normal}$$

Using the data from Table 3.1.1 the Z value is computed for comparison of each level. These results are in Table 3.5.1 below.

Table 3.5.1

Levels Compared	Difference of Means	Standard Deviation	Z	P-Value
1 vs. 2	(191)	102	(1.87)	**0.031**
2 vs. 3	(16)	114	(0.14)	0.446
3 vs. 4	(214)	121	(1.77)	**0.038**

In this example, the hypothesis yields the same class plan as the method of this paper ("1, 2-3, 4"). Level 1 is separated from Level 2 because the difference is significant. Likewise, Level 3 is separated from Level 4. On the other hand, Levels 2 and 3 are combined because their difference is not significant.

[8] Salam suggested using his method on frequency where frequency was Poisson and using a particular formula proved by him. The example of this paper is a pure premium, so the Standard Normal test is used.

3.6 Analysis of the Methods

While Salam's hypothesis testing method and the method of this paper yield the same results in this simple example, they do not always yield the same results. This is because each method has a different goal. The goal of Salam's hypothesis testing method is to separate levels that are statistically significant. The goal of this paper's method is to maximize explained variance while taking into account concerns of credibility.

Explained variance of credibility-weighted means, "Score," measures the squared difference between each class and the overall book mean. Score is a function of three factors:

(1) The difference between the class means;

(2) The number of classes; and

(3) The credibility of each class.

An increase in any of these factors will also increase Score. This can be seen by expressing Score as:

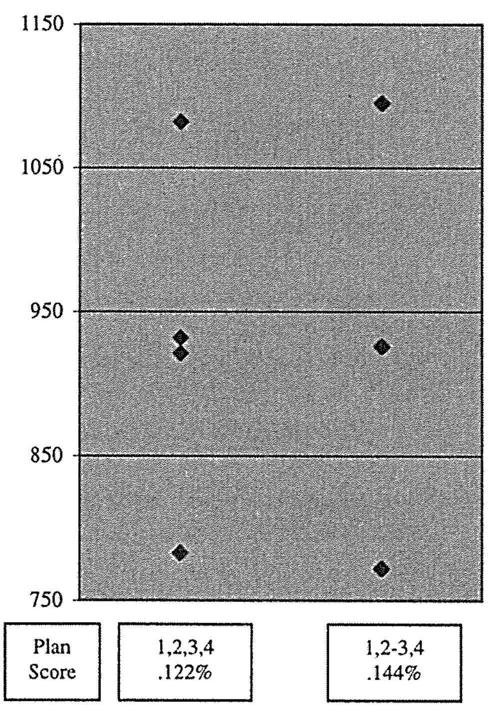
$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{Credibility}_i)\text{^}2 * (\text{BookMean} - \text{ClassMean}_i)\text{^}2}{\text{Constant}}$$

(The proof is in Appendix B.)

As Factor 1 (the difference between the class means) increases, by definition, Score will increase. As Factor 2 (the number of classes) increases, Score increases because an increase in number of classes results in a greater square of the difference of class means. Factor 3 (the credibility of each class) also increases Score, as seen by the formula above.

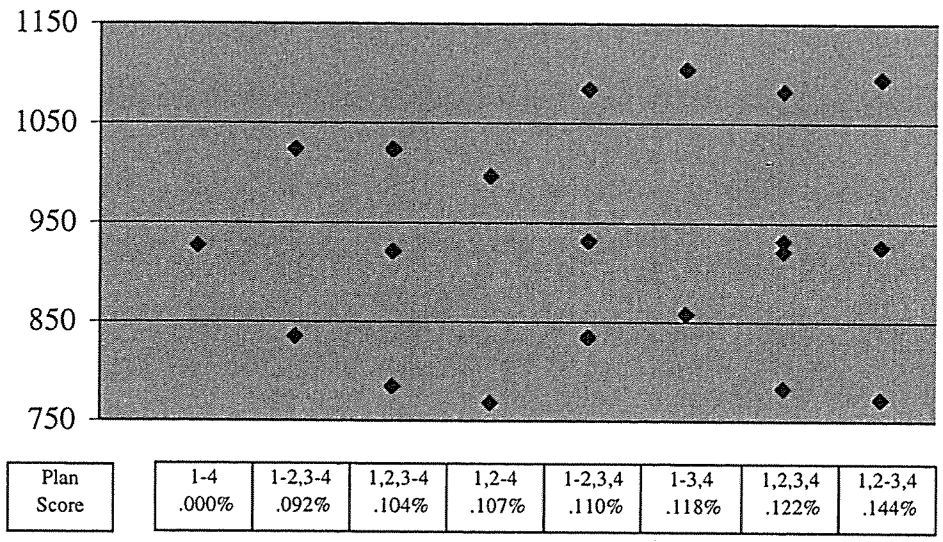
The key to the method of this paper is that although an increase in the number of classes increases Score via Factor 2, an increase in number of classes will also decrease credibility and thus decrease Score via Factor 3. Levels should be combined into one class or split into separate classes based on whether the positive effect from Factor 2 is greater than the negative effect from Factor 3. This can be illustrated through the example of this section. From Tables 3.2.4 and 3.2.1, the average credibility of the consolidated class plan is 81% while the non-consolidated class plan has an average credibility of 70%. In this example, this drop in credibility decreases Score more than the increase from a greater number of classes.

Chart 3.1: Credibility-weighted Class Means



The method can also be illustrated graphically. Chart 3.1 shows the credibility-weighted class means. While plan "1, 2, 3, 4" has more classes, plan "1, 2-3, 4" has a greater spread. In this example, the method of this paper determined that plan "1, 2-3, 4" had a greater variance and thus a greater ability to provide price differentiation. Chart 3.2, below, shows the credibility-weighted class means for each class plan tested.

Chart 3.2: Credibility-weighted Class Means



This paper's method is a natural choice for the determination of class plans. If the company's goal is to provide maximum price differentiation through risk classification, then the company will divide its book into an ever-increasing number of classes. As class size diminishes so will credibility. As credibility decreases, the company will choose to credibility-weigh each class against the entire book. At some point, the drop in credibility actually causes the price differentiation to decrease as each class's mean is overpowered by the complement of credibility. The method in this paper yields the point at which further subdivision will actually decrease price differentiation, as defined by the variance of class means. Therefore, this paper's method creates the maximum amount of price differentiation, given that a company credibility-weighs its small classes.

4. SIMULATION OF A SIMPLE EXAMPLE

As stated earlier, this paper's method and hypothesis testing (Salam) will not always yield the same results. A simulation was done to compare the results of the two methods.

Let the losses follow a binomial distribution for frequency and a gamma distribution for severity. Let the levels be distributed according to Table 4.1 and impose a deductible of \$5,000 and a limit of \$100,000.

Table 4.1

	Frequency	Severity		E(Loss)	Exposures
Level	P(Accident)	Alpha	Beta		
1	0.06	10	1000	600	1,000
2	0.08	10	1100	880	1,000
3	0.08	10	1100	880	1,000
4	0.12	10	1200	1440	1,000

A simulation was performed. 1,000 exposures and policy counts were generated per level. Using the method of this paper, each possible class plan (1, 2-3-4; 1-2, 3-4; 1-2-3, 4; 1-2, 3, 4; 1, 2-3, 4; 1, 2, 3-4; 1, 2, 3, 4; 1-2-3-4) was scored, and the class plan with the highest score was selected (see Appendix E for the simulation program). Additionally, using hypothesis testing in a manner similar to that suggested by Salam,[9] a class plan was

[9] Salam suggested using his method on frequency where frequency was Poisson and using a particular formula proved by him. For this paper losses were tested using the standard hypothesis testing formula described above. Z > 1.96 was used to test for significance.

selected. This simulation was then repeated one hundred times. Appendix C displays the results. Tables 4.2 and 4.3, below, summarize the results.

Table 4.2 – Results of Simulation

Pivot Table							
Count of Attempt	Hypothesis Method (Salam)						
Turner Method	1-2-3-4	1-1,2-4	1-3,4-4	1-1,2-3,4-4	1-2,3-3,4-4	1,2,3,4	Grand Total
1-2-3-4							0
1-1,2-4							0
1-3,4-4			10				10
1-1,2-3,4-4	1	6	15	40			62
1-2,3-3,4-4	3		7		2		12
1,2,3,4			1	14		1	16
Grand Total	4	6	33	54	2	1	100

Table 4.3 – Summary of Simulation

	Did Turner Method Choose "1, 2-3, 4"?	Did Hypothesis Testing Choose "1, 2-3, 4"?	Do Turner and Hypothesis Testing Agree?
Yes	62	54	53
No	38	46	47

Some may be surprised that the hypothesis method chose the theoretically correct plan of "1, 2-3, 4" in only 54% of the simulations; however, the nature of a 95% confidence interval must be properly understood. Using the hypothesis method, if the two levels have the same underlying mean, the procedure will *not* segment the levels 95% of the time. This test is done between levels 1 and 2; 2 and 3; and 3 and 4. There are three places for this test to go wrong, and the hypothesis method makes no claims about Type II errors—failure to segment when segmentation should occur. The hypothesis method is a negative method, only providing segmentation between any two levels when, given its underlying assumptions, a segmentation would be warranted 95 of 100 times. The hypothesis method is similar to the legal standard of "innocent until proven guilty," except it is "no segmentation until segmentation is proven beyond a reasonable doubt." Therefore, this method prefers less segmentation, a potential disadvantage for a company seeking to maximize price differentiation.

The method of this paper yields the highest explained variance of the credibility-weighted means. In this particular example, it selected the theoretically correct class plan in 62% of the simulations, a 15% improvement over the hypothesis testing method. Also note that in four of the simulations the hypothesis method chose "1-2-3-4" (no segmentation) and in thirty-nine of the simulations chose only two classes. In contrast, the method of this paper never chose "1-2-3-4" (no segmentation) and chose a plan with two classes in only ten of the simulations. In this example, this paper's method preferred more segmentation, a potential advantage for a company that is seeking to maximize price differentiation.

The hypothesis method and the method of this paper are useful under different circumstances. The hypothesis method is valuable when societal or business demands create a strong bias in favor of the null hypothesis, which is no segmentation. A recent example of this is the use of credit scores to predict personal automobile accidents. Government bodies were disturbed that people with low credit scores would be charged higher premiums. In this instance, the hypothesis method would be of value because society would demand proof beyond a reasonable doubt before it would allow credit scores to determine automobile premiums. On the other hand, the method of this paper is particularly valuable when considering rating variables that are universally accepted. The use of radius of operation in commercial automobile exemplifies a relatively uncontroversial rating variable. With this variable, the question posed to the actuary will not be *whether* to use this variable but how many partitions should be made.

5. HOW TO PERFORM THIS METHOD ON YOUR DATA—A COMPLEX EXAMPLE

5.1 Introduction to the Example

Suppose a company began writing a specialty product a few years ago at a set rate. Now that the company has a few years of business, it wishes to establish underwriting classes based on various characteristics it has tracked. Principally, it has monitored location, radius of operation, and whether the business is owner-operated. The task is to find the best classification system for price segmentation.

5.2 Data Preparation

Ultimately the data must be summarized by every factor to be included in the study. The fields that must be summed are exposures, policy counts, losses per policy, and the square of losses per policy divided by exposures per policy.

If using a transactional database, the data must first be summarized by policy id. For a summary database, the records must contain policy level information; otherwise, a preceding database must be accessed to obtain policy level information.

The database must contain at the policy level each factor to be included in the study along with exposures, policy counts (both exposures and policy counts may be "1" per policy), and losses. Then one more field is added for each policy and is set equal to the square of that policy's losses divided by that policy's exposures. This field is entitled "LossesSquared."

At this point the data is formatted something like Table 5.2.1:

Table 5.2.1

Policy ID	Location	Radius	Owner Operated	Counts	Exposures	Losses	LossesSquared
77854A5	Rural	Over 10 miles	Yes	1	1	0	0
77943A5	Suburban	Less than 10 miles	Yes	1	2	3,000	4,500,000
78949A6	Urban	Over 10 miles	No	1	1	0	0
78951A6	Urban	Less than 10 miles	Yes	1	1	4,000	16,000,000
***	***	***	***	***	***	***	***

Now that losses have been squared per policy, policy level data is no longer needed. The next step is to summarize by all of the factors in the study and sum exposures, policy counts (which may be the same as exposures), losses, and losses per policy squared. At this point the data should look something like Table 5.2.2:

Table 5.2.2

Location	Radius	Owner Operated	Polcounts	Exposures	Losses	LossesSquared
City	Less than 10 miles	Yes	1,250	1,563	1,092,799	13,345,632,591
Suburban	Less than 10 miles	Yes	1,535	1,919	733,418	8,147,186,304
Rural	Less than 10 miles	Yes	1,014	1,268	593,773	6,724,286,005
City	Over 10 Miles	Yes	207	259	295,748	3,479,777,226
Suburban	Over 10 Miles	Yes	1,272	1,591	886,662	10,558,821,934
Rural	Over 10 Miles	Yes	810	1,013	427,466	4,777,629,538
City	Less than 10 miles	No	2,412	3,016	3,443,920	39,503,485,208
Suburban	Less than 10 miles	No	1,232	1,541	1,687,092	19,948,652,768
Rural	Less than 10 miles	No	1,405	1,757	2,174,428	26,350,261,440
City	Over 10 Miles	No	2,105	2,632	10,316,789	135,036,132,799
Suburban	Over 10 Miles	No	1,890	2,363	3,088,961	35,940,106,619
Rural	Over 10 Miles	No	430	538	871,625	11,051,938,151

5.3 Sort

Next the data must be sorted from "lowest risk" to "highest risk." To do this, losses are simply sorted based on losses per exposure.[10] Table 5.3.1 shows the data from this example sorted from lowest risk to highest risk.

Table 5.3.1

Location	Radius	Owner Operated	Polcounts	Exposures	Losses	LossesSquared	Losses / Exposure
Suburban	Less than 10 miles	Yes	1,535	1,919	733,418	8,147,186,304	382
Rural	Over 10 Miles	Yes	810	1,013	427,466	4,777,629,538	422
Rural	Less than 10 miles	Yes	1,014	1,268	593,773	6,724,286,005	468
Suburban	Over 10 Miles	Yes	1,272	1,591	886,662	10,558,821,934	557
City	Less than 10 miles	Yes	1,250	1,563	1,092,799	13,345,632,591	699
Suburban	Less than 10 miles	No	1,232	1,541	1,687,092	19,948,652,768	1,095
City	Less than 10 miles	No	2,412	3,016	3,443,920	39,503,485,208	1,142
City	Over 10 Miles	Yes	207	259	295,748	3,479,777,226	1,142
Rural	Less than 10 miles	No	1,405	1,757	2,174,428	26,350,261,440	1,238
Suburban	Over 10 Miles	No	1,890	2,363	3,088,961	35,940,106,619	1,307
Rural	Over 10 Miles	No	430	538	871,625	11,051,938,151	1,620
City	Over 10 Miles	No	2,105	2,632	10,316,789	135,036,132,799	3,920

Once the risks are ranked, the verbal identifiers are discarded in favor of reliance simply on the rankings. A key is used to match the rankings back to the verbal identifiers. Table 5.3.2 shows the key while Table 5.3.3 shows data by rank. Table 5.3.3 is ready to be used as input for the computer program in Appendix D.

[10] If the resulting rank order produces results that are out of line with business or regulatory constraints, a deviation from the accepted sort is warranted. The levels are placed in the order that would be acceptable from a regulatory or business standpoint. This will produce suboptimal results, but will produce the best results given the "constraints." For example, in Table 6.3.1, the level "Rural, Less than 10 Miles, and Not Owner Operated" has a lot of losses per exposure. It may be that everyone believes this is a fluke and considers this level to be a good risk. If it would be politically impossible to put this level into a risky underwriting class, the actuary may need to arbitrarily place this row higher up in the table. In the Simple Example, this step was skipped altogether, because it was understood that the risk was increasing as the level increased.

Table 5.3.2

Index1	Location	Radius	Owner Operated
1	Suburban	Less than 10 miles	Yes
2	Rural	Over 10 Miles	Yes
3	Rural	Less than 10 miles	Yes
4	Suburban	Over 10 Miles	Yes
5	City	Less than 10 miles	Yes
6	Suburban	Less than 10 miles	No
7	City	Less than 10 miles	No
8	City	Over 10 Miles	Yes
9	Rural	Less than 10 miles	No
10	Suburban	Over 10 Miles	No
11	Rural	Over 10 Miles	No
12	City	Over 10 Miles	No

Table 5.3.3

Index1	Polcounts	Exposures	Losses	LossesSquared
1	1,535	1,919	733,418	8,147,186,304
2	810	1,013	427,466	4,777,629,538
3	1,014	1,268	593,773	6,724,286,005
4	1,272	1,591	886,662	10,558,821,934
5	1,250	1,563	1,092,799	13,345,632,591
6	1,232	1,541	1,687,092	19,948,652,768
7	2,412	3,016	3,443,920	39,503,485,208
8	207	259	295,748	3,479,777,226
9	1,405	1,757	2,174,428	26,350,261,440
10	1,890	2,363	3,088,961	35,940,106,619
11	430	538	871,625	11,051,938,151
12	2,105	2,632	10,316,789	135,036,132,799

5.4 Produce a Score for Each Possible Class Plan

The data shown in Table 5.3.3 is imported into an Access Database. The program in the Appendix is then run after it has been modified to meet the needs of the particular study. The program iterates over every possible combination where the levels follow the rank established in the field Index1. Because Index1 is ranked from least to most risky, the program does not need to iterate through combinations that do not follow the ranking. For example, the program needs not attempt to create a class plan such as "Class 1 = 1,2,11,12, and Class 2=3,4,5,6,7,8,9,10" because there is no possibility that such a class plan could produce the highest Score.

Table 5.4.1, below, shows the results of the computer program. All 2,048 possible class plans were analyzed. In order to conserve space, only the best five and worst five class plans are shown.

Table 5.4.1

Class Plans	Score
1-4, 5-5,6-8,9-10,11-11,12-12	8.10%
1-3, 4-4,5-5,6-8,9-10,11-12	8.10%
1-4, 5-5,6-8,9-9,10-10,11-12	8.10%
1-3, 4-4,5-5,6-8,9-9,10-12	8.10%
1-3, 4-5,6-8,9-10,11-11,12-12	8.10%
***	***
The intermediate 2,038 Class Plans are omitted.	
***	***
1-1, 2-2,3-3,4-12	1.49%
1-2, 3-12	1.02%
1-1, 2-2,3-12	1.00%
1-1, 2-12	0.64%
1-12 (no segmentation)	0.00%

The worst class plan is "1-12" (no segmentation) while the best class plan is "1-4, 5-5,6-8,9-10,11-11,12-12." Note that the better class plans have over five times the explanatory power of the poorer plans. This demonstrates the advantage of using a technique such as the one in this paper. Also note that the better class plans have six classes, not twelve. This shows that dividing the book into all available levels does not necessarily yield the highest price segmentation after considering credibility. Finally, note that the best five plans had similar Scores. Any of the five would probably make acceptable class plans, and a selection could be made from these based on business convenience.

Assuming that the class plan of "1-4, 5-5,6-8,9-10,11-11,12-12" (the plan with the highest Score) is selected, the next step is to develop the underwriting description of the class plan. According to the key (Table 5.3.2), the underwriting guidelines should be as in Table 5.4.2 below.

Table 5.4.2

Index	Underwriting Class	Location	Radius	Owner Operated
1	A	Less than 10 miles	Suburban	Yes
2	A	Over 10 Miles	Rural	Yes
3	A	Less than 10 miles	Rural	Yes
4	A	Over 10 Miles	Suburban	Yes
5	B	Less than 10 miles	City	Yes
6	C	Less than 10 miles	Suburban	No
7	C	Less than 10 miles	City	No
8	C	Over 10 Miles	City	Yes
9	D	Less than 10 miles	Rural	No
10	D	Over 10 Miles	Suburban	No
11	E	Over 10 Miles	Rural	No
12	F	Over 10 Miles	City	No

This can be explained in an underwriting manual as Table 5.4.3

Table 5.4.3

Underwriting Class	Guidelines
A	Owner Operated-Suburban or Rural
B	Owner Operated-City-Less than 10 Miles
C	Owner Operated-City-More than 10 Miles; or, Not Owner Operated-City or Suburban-Less than 10 Miles
D	Not Owner Operated-Rural-Less than 10 Miles or Suburban-Over 10 Miles
E	Not Owner Operated-Rural-Over 10 Miles
F	Not Owner Operated-City-Over 10 Miles

7. CONCLUSION

Actuaries are called upon to create risk classification systems. In an effort to skim the cream, companies often desire plans with the maximum price differentiation. Common sense dictates that at some point, too much refinement is counterproductive. This paper introduces one method for determining the point at which further segmentation defeats the purpose of price segmentation. A simulation of this method versus hypothesis testing is shown, and the results are encouraging. Finally, this paper demonstrates how one can perform such an analysis on his or her data and provides sample code to assist in the process.

APPENDIX A

Definitions:[11]
Number of Classes = Number of underwriting classes

$PolicyCount_i$ = Policy count in underwriting class i

Exposures = All exposures

$Exposures_i$ = Exposures per class i

$Exposures_{ij}$ = Exposures per policy j in class i

Losses = All losses

$Losses_i$ = Losses per class i

$Losses_{ij}$ = Losses per policy j in class i

$LossesSquared_{ij}$ = (Losses per policy j in class i)^2 / $Exposures_{ij}$

$LossesSquared_i$ = $LossesSquared_{ij}$ per class i

Formulas
To maximize the credible explanatory power, maximize "Score" below.

$$\mathbf{C} = \Sigma_i\Sigma_j(Losses_{ij}\text{^}2 / Exposures_{ij}) = \Sigma_i\Sigma_j(LossesSquared_{ij}) = \Sigma_i (LossesSquared_i)$$

$$\mathbf{V} = \{\frac{1}{\Sigma(PolicyCounts_i - 1)}\} * \{(C - (\Sigma \frac{Losses_i \text{^}2}{Exposures_i}))\}$$

$$\mathbf{A} = \frac{(\Sigma\, W_i) - V * (\text{Number of Classes - 1})}{(\Sigma\, Exposures_i) - \frac{\Sigma\, Exposures_i \text{^}2}{\Sigma\, Exposures_i}}$$

$$\mathbf{K} = V/A$$

$\mathbf{Credibility_i}$ = [$Exposures_i$ / ($Exposures_i$ + K)] {From 5.76 at Klugman [3]}

$$\mathbf{BookMean} = (\sum_{ij} Losses_{ij}) / (\sum_{ij} Exposures_{ij})$$

$\mathbf{Cred.Wtd.ClassMean_i}$ = $Credibility_i$ * $Losses_i$ / $Exposures_i$ + (1 - $Credibility_i$)* BookMean

$$\mathbf{Score} = \frac{\sum_i Exposures_i \times (BookMean - Cred.Wtd.ClassMean_i)\text{^}2}{(\sum_i LossesSquared_i) - (Exposures \times [BookMean]\text{^}2)}$$

---

[11] "^2" means "squared"

APPENDIX B

$$\textbf{Cred.Wtd.ClassMean}_i = \text{Credibility}_i * \text{Losses}_i / \text{Exposures}_i + (1 - \text{Credibility}_i) * \text{BookMean} \qquad \{1\}$$

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{BookMean} - \text{Cred.Wtd.ClassMean}_i)\text{^}2}{(\sum_i \text{LossesSquared}_i) - (\text{Exposures} \times [\text{BookMean}]\text{^}2)} \qquad \{2\}$$

The formula below is produced by substituting {1} into {2}:

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{BookMean} - \{\text{Credibility}_i * \text{Losses}_i / \text{Exposures}_i + (1 - \text{Credibility}_i) * \text{BookMean}\})\text{^}2}{(\sum_i \text{LossesSquared}_i) - (\text{Exposures} \times [\text{BookMean}]\text{^}2)}$$

This can be simplified to:

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (-\text{Credibility}_i * \text{Losses}_i / \text{Exposures}_i + \text{Credibility}_i * \text{BookMean})\text{^}2}{(\sum_i \text{LossesSquared}_i) - (\text{Exposures} \times [\text{BookMean}]\text{^}2)}$$

Which further simplifies to:

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{Credibility}_i)\text{^}2 * (\text{BookMean} - \text{Losses}_i / \text{Exposures}_i)\text{^}2}{(\sum_i \text{LossesSquared}_i) - (\text{Exposures} \times [\text{BookMean}]\text{^}2)}$$

By noting that the denominator is a constant, the formula becomes:

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{Credibility}_i)\text{^}2 * (\text{BookMean} - \text{Losses}_i / \text{Exposures}_i)\text{^}2}{\text{Constant}}$$

Finally, by noting that "$\text{Losses}_i / \text{Exposures}_i$" is merely the mean for each class:

$$\text{Score} = \frac{\sum_i \text{Exposures}_i \times (\text{Credibility}_i)\text{^}2 * (\text{BookMean} - \text{ClassMean}_i)\text{^}2}{\text{Constant}}$$

APPENDIX C

Attempt	Turner Method	Hypothesis Method (Salam)	Turner "Correct"?	Hypothesis Testing "Correct"?	Do Score and Turner Agree?
1	1,2,3,4	1-1,2-3,4-4	No	Yes	No
2	1-2,3-3,4-4	No Segmentation	No	No	No
3	1,2,3,4	1-1,2-3,4-4	No	Yes	No
4	1-2,3-3,4-4	1-3,4-4	No	No	No
5	1-2,3-3,4-4	1-3,4-4	No	No	No
6	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
7	1-1,2-3,4-4	1-3,4-4	Yes	No	No
8	1-1,2-3,4-4	1-3,4-4	Yes	No	No
9	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
10	1-1,2-3,4-4	1-3,4-4	Yes	No	No
11	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
12	1-2,3-3,4-4	1-2,3-3,4-4	No	No	Yes
13	1,2,3,4	1,2,3,4	No	No	Yes
14	1-1,2-3,4-4	1-3,4-4	Yes	No	No
15	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
16	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
17	1-1,2-3,4-4	1-3,4-4	Yes	No	No
18	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
19	1-2,3-3,4-4	1-3,4-4	No	No	No
20	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
21	1-3,4-4	1-3,4-4	No	No	Yes
22	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
23	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
24	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
25	1-1,2-3,4-4	1-3,4-4	Yes	No	No
26	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
27	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
28	1-1,2-3,4-4	1-3,4-4	Yes	No	No
29	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
30	1-3,4-4	1-3,4-4	No	No	Yes
31	1-3,4-4	1-3,4-4	No	No	Yes
32	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
33	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
34	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
35	1,2,3,4	1-1,2-3,4-4	No	Yes	No
36	1-3,4-4	1-3,4-4	No	No	Yes
37	1,2,3,4	1-1,2-3,4-4	No	Yes	No
38	1-2,3-3,4-4	No Segmentation	No	No	No
39	1-3,4-4	1-3,4-4	No	No	Yes
40	1-2,3-3,4-4	1-3,4-4	No	No	No
41	1-3,4-4	1-3,4-4	No	No	Yes
42	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
43	1-1,2-3,4-4	1-1,2-4	Yes	No	No
44	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
45	1,2,3,4	1-1,2-3,4-4	No	Yes	No
46	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
47	1-1,2-3,4-4	1-1,2-4	Yes	No	No
48	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
49	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
50	1-1,2-3,4-4	1-1,2-4	Yes	No	No
51	1-1,2-3,4-4	1-3,4-4	Yes	No	No
52	1,2,3,4	1-1,2-3,4-4	No	Yes	No
53	1-3,4-4	1-3,4-4	No	No	Yes
54	1,2,3,4	1-1,2-3,4-4	No	Yes	No
55	1-1,2-3,4-4	1-3,4-4	Yes	No	No
56	1,2,3,4	1-1,2-3,4-4	No	Yes	No
57	1-1,2-3,4-4	No Segmentation	Yes	No	No

58	1-1,2-3,4-4	1-1,2-4	Yes	No	No
59	1-2,3-3,4-4	1-3,4-4	No	No	No
60	1-2,3-3,4-4	1-3,4-4	No	No	No
61	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
62	1-2,3-3,4-4	No Segmentation	No	No	No
63	1-2,3-3,4-4	1-3,4-4	No	No	No
64	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
65	1,2,3,4	1-1,2-3,4-4	No	Yes	No
66	1-1,2-3,4-4	1-3,4-4	Yes	No	No
67	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
68	1-1,2-3,4-4	1-1,2-4	Yes	No	No
69	1-3,4-4	1-3,4-4	No	No	Yes
70	1,2,3,4	1-3,4-4	No	No	No
71	1-1,2-3,4-4	1-3,4-4	Yes	No	No
72	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
73	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
74	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
75	1-3,4-4	1-3,4-4	No	No	Yes
76	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
77	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
78	1-1,2-3,4-4	1-1,2-4	Yes	No	No
79	1,2,3,4	1-1,2-3,4-4	No	Yes	No
80	1,2,3,4	1-1,2-3,4-4	No	Yes	No
81	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
82	1-3,4-4	1-3,4-4	No	No	Yes
83	1-1,2-3,4-4	1-3,4-4	Yes	No	No
84	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
85	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
86	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
87	1-2,3-3,4-4	1-2,3-3,4-4	No	No	Yes
88	1-1,2-3,4-4	1-3,4-4	Yes	No	No
89	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
90	1,2,3,4	1-1,2-3,4-4	No	Yes	No
91	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
92	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
93	1-1,2-3,4-4	1-3,4-4	Yes	No	No
94	1-1,2-3,4-4	1-3,4-4	Yes	No	No
95	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
96	1,2,3,4	1-1,2-3,4-4	No	Yes	No
97	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
98	1,2,3,4	1-1,2-3,4-4	No	Yes	No
99	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes
100	1-1,2-3,4-4	1-1,2-3,4-4	Yes	Yes	Yes

APPENDIX D (Computer Program – Up to 12 Possible Classes)

Below are the steps to running the program, followed by the Program:

1) Paste your data, formatted like Table 5.3.3, into a Microsoft Access Database. Call the Table "Data."
2) Close Microsoft Access. (To avoid sharing violations)
3) Open Microsoft Excel. Click Tools -> Macros -> Visual Basic Editor. Once you are within the Visual Basic Editor, click "Insert Module."
4) Make sure the module has nothing written in it. (Sometimes Microsoft Excel inserts in default settings.)
5) Paste the program into the module.
6) Click Tools-> References -> Microsoft DAO 3.6 Object Library. Put a checkmark next to it and click "ok."
7) Make modifications A and B within the program.
8) Run the program by placing your cursor within the subroutine called "PerformAnalysis," and click Run -> Run Sub/User Form.
9) Wait for the message box which says "Completed."
10) Open the Microsoft Access Database and check out the table "Iterations." It shows the score for each iteration, in the order it was iterated. Copy this table into Excel and sort by score descending. An iteration with a high Score should be selected.

The program is below:

```
Dim cols As New Collection
  Dim colTiers As New Collection     ' Create a Collection object.
  Dim SQL, Guy, TaBler As String
  Dim namer
  Dim dber
  Dim Alpha, Beta, P, NumberToGenerate As Double
  Dim Exposures, Losses, LossesSquared As Double
  Dim Deductible, Limit As Double
  Dim Index1 As Double
  Dim NumberofFactors As Double
  Dim countExcel As Integer
  Dim iter1


Sub PerformAnalysis()
' MODIFICATION A - CHANGE THE NAME OF THE DATABASE BELOW TO MATCH YOUR DATABASE.
    Set dber = OpenDatabase("C:\Documents and Settings\Benjamin Turner\My
Documents\Turner Score\aug23c.mdb")
Call PerformScoreAnalysis
MsgBox "Completed."
End Sub
Sub PerformScoreAnalysis()
    SQL = "create table Iterations (Title text, Score double)"
dber.Execute (SQL)
     SQL = "select * into standard from data"
dber.Execute (SQL)
     SQL = "alter table standard add column index2 integer"
dber.Execute (SQL)
' MODIFICATION B - CHANGE THE NUMBER OF FACTORS TO THE NUMBER OF POSSIBLE CLASSES.
  ' IN THE SIMPLE EXAMPLE THIS WAS 4 AND IN THE COMPLEX EXAMPLE THIS WAS 12.
  ' THIS PROGRAM CAN HANDLE 2 TO 12 POSSIBLE POTENTIAL CLASSES.
  ' SOMEONE FAMILIAR WITH EXCEL MACROS COULD MODIFY THIS PROGRAM TO DO MORE THAN 12
CLASSES

NumberofFactors = 4
Call OneClass
If NumberofFactors > 2 Then Call TwoClasses
```

```
 If NumberofFactors > 3 Then Call ThreeClasses
 If NumberofFactors > 4 Then Call FourClasses
 If NumberofFactors > 5 Then Call FiveClasses
 If NumberofFactors > 6 Then Call SixClasses
 If NumberofFactors > 7 Then Call SevenClasses
 If NumberofFactors > 8 Then Call EightClasses
 If NumberofFactors > 9 Then Call NineClasses
 If NumberofFactors > 10 Then Call TenClasses
 If NumberofFactors > 11 Then Call ElevenClasses
 Call AllClasses
 End Sub

 Sub OneClass()
 'Only one class
     SQL = "insert into iterations (title, score) values ('No
 Segmentation',0.000000000000000001)"
 dber.Execute (SQL)
 'End only one class
 End Sub

 Sub AllClasses()
 'All classes


     SQL = "update standard set index2 = index1"
     namer = "1"
     For i = 2 To NumberofFactors
       namer = namer & "," & i
     Next

  dber.Execute (SQL)
  Call GetScore
  'End All classes
 End Sub

 Sub TwoClasses()
 'Two Classes
 For i1 = 1 To NumberofFactors - 1
     SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
  dber.Execute (SQL)
     SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " &
 NumberofFactors
  dber.Execute (SQL)
     namer = "1-" & i1 & ", " & i1 + 1 & "-" & NumberofFactors
  Call GetScore
 Next
 'End Two classes

 End Sub

 Sub ThreeClasses()
 'Three Classes
 For i1 = 1 To NumberofFactors - 2
 For i2 = i1 + 1 To NumberofFactors - 1

     SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
  dber.Execute (SQL)
     SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
  dber.Execute (SQL)
     SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " &
 NumberofFactors
  dber.Execute (SQL)

     namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & NumberofFactors
  Call GetScore
 Next
 Next
 ' end three classes
 End Sub


 Sub FourClasses()
 'Four Classes
 For i1 = 1 To NumberofFactors - 3
 For i2 = i1 + 1 To NumberofFactors - 2
 For i3 = i2 + 1 To NumberofFactors - 1

     SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
  dber.Execute (SQL)
     SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
  dber.Execute (SQL)
```

```
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & NumberofFactors
 Call GetScore
Next
Next
Next
' end Four classes

End Sub

Sub FiveClasses()
'Five Classes
For i1 = 1 To NumberofFactors - 4
For i2 = i1 + 1 To NumberofFactors - 3
For i3 = i2 + 1 To NumberofFactors - 2
For i4 = i3 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4 & "," & i4 + 1 & "-" & NumberofFactors

 Call GetScore
Next
Next
Next
Next
' end Five classes

End Sub
Sub SixClasses()
For i1 = 1 To NumberofFactors - 5
For i2 = i1 + 1 To NumberofFactors - 4
For i3 = i2 + 1 To NumberofFactors - 3
For i4 = i3 + 1 To NumberofFactors - 2
For i5 = i4 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5 & "," & i5 + 1 & "-" & NumberofFactors

 Call GetScore
Next
Next
Next
Next
Next
End Sub
Sub SevenClasses()
For i1 = 1 To NumberofFactors - 6
For i2 = i1 + 1 To NumberofFactors - 5
```

```
For i3 = i2 + 1 To NumberofFactors - 4
For i4 = i3 + 1 To NumberofFactors - 3
For i5 = i4 + 1 To NumberofFactors - 2
For i6 = i5 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " & i6
 dber.Execute (SQL)

    SQL = "update standard set index2 = 7 where index1 between " & i6 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5
    namer = namer & "," & i5 + 1 & "-" & NumberofFactors
 Call GetScore
Next
Next
Next
Next
Next
Next
End Sub
Sub EightClasses()
For i1 = 1 To NumberofFactors - 7
For i2 = i1 + 1 To NumberofFactors - 6
For i3 = i2 + 1 To NumberofFactors - 5
For i4 = i3 + 1 To NumberofFactors - 4
For i5 = i4 + 1 To NumberofFactors - 3
For i6 = i5 + 1 To NumberofFactors - 2
For i7 = i6 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " & i6
 dber.Execute (SQL)
    SQL = "update standard set index2 = 7 where index1 between " & i6 + 1 & " and " & i7
 dber.Execute (SQL)

    SQL = "update standard set index2 = 8 where index1 between " & i7 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5
    namer = namer & "," & i5 + 1 & "-" & NumberofFactors
 Call GetScore
Next
Next
Next
Next
Next
Next
Next
End Sub
Sub NineClasses()
For i1 = 1 To NumberofFactors - 8
For i2 = i1 + 1 To NumberofFactors - 7
For i3 = i2 + 1 To NumberofFactors - 6
For i4 = i3 + 1 To NumberofFactors - 5
For i5 = i4 + 1 To NumberofFactors - 4
```

```
For i6 = i5 + 1 To NumberofFactors - 3
For i7 = i6 + 1 To NumberofFactors - 2
For i8 = i7 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " & i6
 dber.Execute (SQL)
    SQL = "update standard set index2 = 7 where index1 between " & i6 + 1 & " and " & i7
 dber.Execute (SQL)
    SQL = "update standard set index2 = 8 where index1 between " & i7 + 1 & " and " & i8
 dber.Execute (SQL)

    SQL = "update standard set index2 = 9 where index1 between " & i8 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5
    namer = namer & "," & i5 + 1 & "-" & NumberofFactors
 Call GetScore
Next
Next
Next
Next
Next
Next
Next
Next
End Sub
Sub TenClasses()
For i1 = 1 To NumberofFactors - 9
For i2 = i1 + 1 To NumberofFactors - 8
For i3 = i2 + 1 To NumberofFactors - 7
For i4 = i3 + 1 To NumberofFactors - 6
For i5 = i4 + 1 To NumberofFactors - 5
For i6 = i5 + 1 To NumberofFactors - 4
For i7 = i6 + 1 To NumberofFactors - 3
For i8 = i7 + 1 To NumberofFactors - 2
For i9 = i8 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " & i6
 dber.Execute (SQL)
    SQL = "update standard set index2 = 7 where index1 between " & i6 + 1 & " and " & i7
 dber.Execute (SQL)
    SQL = "update standard set index2 = 8 where index1 between " & i7 + 1 & " and " & i8
 dber.Execute (SQL)
    SQL = "update standard set index2 = 9 where index1 between " & i8 + 1 & " and " & i9
 dber.Execute (SQL)

    SQL = "update standard set index2 = 10 where index1 between " & i9 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5
    namer = namer & "," & i5 + 1 & "-" & NumberofFactors
 Call GetScore
Next
Next
Next
```

```
Next
Next
Next
Next
Next
Next
End Sub

Sub ElevenClasses()
For i1 = 1 To NumberofFactors - 10
For i2 = i1 + 1 To NumberofFactors - 9
For i3 = i2 + 1 To NumberofFactors - 8
For i4 = i3 + 1 To NumberofFactors - 7
For i5 = i4 + 1 To NumberofFactors - 6
For i6 = i5 + 1 To NumberofFactors - 5
For i7 = i6 + 1 To NumberofFactors - 4
For i8 = i7 + 1 To NumberofFactors - 3
For i9 = i8 + 1 To NumberofFactors - 2
For i10 = i9 + 1 To NumberofFactors - 1

    SQL = "update standard set index2 = 1 where index1 between 1 and " & i1
 dber.Execute (SQL)
    SQL = "update standard set index2 = 2 where index1 between " & i1 + 1 & " and " & i2
 dber.Execute (SQL)
    SQL = "update standard set index2 = 3 where index1 between " & i2 + 1 & " and " & i3
 dber.Execute (SQL)
    SQL = "update standard set index2 = 4 where index1 between " & i3 + 1 & " and " & i4
 dber.Execute (SQL)
    SQL = "update standard set index2 = 5 where index1 between " & i4 + 1 & " and " & i5
 dber.Execute (SQL)
    SQL = "update standard set index2 = 6 where index1 between " & i5 + 1 & " and " & i6
 dber.Execute (SQL)
    SQL = "update standard set index2 = 7 where index1 between " & i6 + 1 & " and " & i7
 dber.Execute (SQL)
    SQL = "update standard set index2 = 8 where index1 between " & i7 + 1 & " and " & i8
 dber.Execute (SQL)
    SQL = "update standard set index2 = 9 where index1 between " & i8 + 1 & " and " & i9
 dber.Execute (SQL)
    SQL = "update standard set index2 = 10 where index1 between " & i9 + 1 & " and " &
i10
 dber.Execute (SQL)

    SQL = "update standard set index2 = 11 where index1 between " & i10 + 1 & " and " &
NumberofFactors
 dber.Execute (SQL)

    namer = "1-" & i1 & ", " & i1 + 1 & "-" & i2 & "," & i2 + 1 & "-" & i3 & "," & i3 + 1
& "-" & i4
    namer = namer & "," & i4 + 1 & "-" & i5
    namer = namer & "," & i5 + 1 & "-" & NumberofFactors
 Call GetScore
Next
Next
Next
Next
Next
Next
Next
Next
Next
Next
End Sub



'Sub GetScore()

TaBler = "Groupings"


    SQL = "select index2, sum(polcounts) as polcountsT, sum(exposures) as exposuresT,"
    SQL = SQL & " sum(losses) as lossesT, sum(lossesSquared) as lossesSquaredT"
    SQL = SQL & " into groupings from standard group by index2"
 dber.Execute (SQL)
SQL = "select * into Work1 from " & TaBler
 dber.Execute (SQL)
```

```
   Dim cols As New Collection
cols.Add Item:="AA", Key:="1"
cols.Add Item:="BB", Key:="2"
cols.Add Item:="CC", Key:="3"
cols.Add Item:="DD", Key:="4"
cols.Add Item:="EE", Key:="5"
cols.Add Item:="FF", Key:="6"
cols.Add Item:="GG", Key:="7"
cols.Add Item:="HH", Key:="8"
cols.Add Item:="II", Key:="9"


For Each Column In cols
  SQL = "alter table Work1"
  SQL = SQL & " add " & Column & " double "
  dber.Execute (SQL)
Next

  SQL = "update work1 set AA = PolCountsT - 1"
dber.Execute (SQL)
  SQL = "update work1 set BB = LossesT*LossesT/ExposuresT"
dber.Execute (SQL)
  SQL = "update work1 set DD = 1"
dber.Execute (SQL)
  SQL = "update work1 set EE = ExposuresT*ExposuresT"
dber.Execute (SQL)

  SQL = "select sum(ExposuresT) as E, sum(LossesT) as L, sum(LossesSquaredT) as C"
  SQL = SQL & ", sum(DD) as DDsm, sum(EE) as EEsm, "
' Set V
  SQL = SQL & " (1/sum(AA))*(sum(LossesSquaredT)- sum(BB)) as v"
  SQL = SQL & " into Work2 from Work1"
dber.Execute (SQL)

  SQL = "update work1, work2 set work1.CC = work1.ExposuresT"
  SQL = SQL & "*(work1.LossesT/work1.ExposuresT - work2.L/work2.E)"
  SQL = SQL & "*(work1.LossesT/work1.ExposuresT - work2.L/work2.E)"
'The last line is to square the parenthetic expression
dber.Execute (SQL)



   SQL = "Alter table work2 add CCsm double"
dber.Execute (SQL)

  SQL = "select sum(work1.CC) As CCsm into work3 from work1"
dber.Execute (SQL)

   SQL = "Alter table work2 add a double"
dber.Execute (SQL)
' "a" is the name of the field

   SQL = "update work2, work3 "
   SQL = SQL & " set work2.a="
   SQL = SQL & "(work3.CCsm-work2.v*(work2.DDsm-1))"
   SQL = SQL & "/(work2.E-(1/work2.E)*work2.EEsm)"

dber.Execute (SQL)

   SQL = "select v/a as k into work4 from work2"
dber.Execute (SQL)


   SQL = "update work1, work4 set work1.ff = work1.ExposuresT/(work1.ExposuresT +
work4.k)"
dber.Execute (SQL)

  SQL = "update work1, work2 set work1.ff = 0 where work2.a < 0"
dber.Execute (SQL)

   SQL = "select sum(lossesT)/sum(exposuresT) as U into work5 from work1"
dber.Execute (SQL)

   SQL = "update work1, work5 set work1.hh= work1.ff*work1.LossesT/work1.ExposuresT    "
   SQL = SQL & " + (1-work1.ff) * work5.u"
dber.Execute (SQL)

   SQL = "update work1, work5 set work1.ii = work1.ExposuresT * "
   SQL = SQL & "(work1.hh - work5.u) * (work1.hh - work5.u)"
dber.Execute (SQL)
```

```
    SQL = "select sum(LossesSquaredT)/sum(ExposuresT) -
(sum(LossesT)/sum(ExposuresT))*(sum(LossesT)/sum(ExposuresT)) "
    SQL = SQL & " as const into Work6 from work1"
dber.Execute (SQL)

   SQL = "select sum(ii)/sum(ExposuresT) as Score into Results from work1"
dber.Execute (SQL)

   SQL = "update Results, work6 set Results.score = Results.score / work6.const"
dber.Execute (SQL)

    SQL = "Alter table Results add title text"
dber.Execute (SQL)

    SQL = "update results set title = '" & namer & "'"
dber.Execute (SQL)

   SQL = "insert into iterations select * from results"
dber.Execute (SQL)


For i = 1 To 6
    SQL = "drop table work" & i
dber.Execute (SQL)
Next i
    SQL = "drop table Results"
dber.Execute (SQL)
    SQL = "drop table " & TaBler

dber.Execute (SQL)
End Sub
```

APPENDIX E (Simulation Computer Program – This is the program that generated the data for Section 4.)

Note. This program is not necessary to perform the analysis. It is provided for those who would like to see the simulation program.

Below are the steps to running the program, followed by the Program:

1) Install the program in Appendix C, but do not place any data into Access.
2) Make sure that you make Modification B, setting "NumberofFactors" equal to 4.
3) Paste this program, Appendix D, below the code from Appendix C. (Make sure that you paste this program in the same module as the program from Appendix C; otherwise is will not run.)
4) Make modification C within the program.
5) Run the program by placing your cursor within the subroutine called "PerformRandomAnalysis," and click Run -> Run Sub/User Form.
6) Wait for the message box that says "Completed".
7) Open the Microsoft Access Database and check out the table "Answers" and "Results Summary."
8) If you wish to compare versus the "Hypothesis Testing" method, use the data outputted to Excel.

The program is below:

```
Sub PerformRandomDataAnalysis()

'MODIFICATION C: CHANGE THE NAME OF THE DATABASE.
    Set dber = OpenDatabase("C:\Documents and Settings\Benjamin Turner\My
Documents\Turner Score\sept1a.mdb")

countExcel = 2

    SQL = "create table Answers (Title text, Score double, Attempt integer)"
dber.Execute (SQL)

'answers2h
    SQL = "create table Answers2h (Title text, Score double, Attempt integer)"
dber.Execute (SQL)
'end answers2h

    Sheets.Add
    Range("A1").Select
    ActiveCell.FormulaR1C1 = "Attempt"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "Group"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "Count"
    Range("D1").Select
    ActiveCell.FormulaR1C1 = "Exposures"
    Range("E1").Select
    ActiveCell.FormulaR1C1 = "Losses"
    Range("F1").Select
    ActiveCell.FormulaR1C1 = "LossesSquared"
    Range("G1").Select
    ActiveCell.FormulaR1C1 = "Variance"
    Range("H1").Select
    ActiveCell.FormulaR1C1 = "Sqrt"
    Range("I1").Select
    ActiveCell.FormulaR1C1 = "E(x)"
    Range("J1").Select
    ActiveCell.FormulaR1C1 = "Z"
    Range("L1").Select
    ActiveCell.FormulaR1C1 = "Normal"
```

```
     Range("M1").Select
     ActiveCell.FormulaR1C1 = "Z Test"


For iter1 = 1 To 100
Call randomNumbers
Call PerformScoreAnalysis

'close out of this iteration.
     SQL = "drop table Data"
dber.Execute (SQL)

     SQL = "drop table Standard"
dber.Execute (SQL)

     SQL = "insert  into answers"
     SQL = SQL & " select * from iterations"
     SQL = SQL & " where score in (select max(score) from iterations)"
dber.Execute (SQL)
     SQL = " update answers set attempt = " & iter1 & " where attempt is null"
dber.Execute (SQL)

'answers2h
     SQL = "insert  into answers2h"
     SQL = SQL & " select * from iterations"
     SQL = SQL & " where score in (select max(score) from iterations)"
dber.Execute (SQL)

     SQL = "insert  into answers2h "
     SQL = SQL & " SELECT *"
     SQL = SQL & " FROM iterations"
     SQL = SQL & " WHERE score in ( select max(score) from iterations"
     SQL = SQL & " where score not in (select max(score) from iterations));"
dber.Execute (SQL)

     SQL = " update answers2h set attempt = " & iter1 & " where attempt is null"
dber.Execute (SQL)
'end answers2h


     SQL = "drop table iterations"
dber.Execute (SQL)
'end iteration
Next iter1

     SQL = "SELECT title, count(title) as total"
     SQL = SQL & " into ResultsSummary FROM Answers"
     SQL = SQL & " group by title"
dber.Execute (SQL)

MsgBox "Completed"
End Sub
Sub randomNumbers()

     SQL = "create table Data (Index1 long, polcounts long, Exposures long, Losses double,
LossesSquared double)"
     dber.Execute (SQL)

     Deductible = 5000
     Limit = 100000
     NumberEquals = 1000

     Alpha = 10
     moder = 1


Index1 = 1
     P = 0.08
     Beta = 900 * moder
     NumberToGenerate = Excel.WorksheetFunction.Round(NumberEquals * 1.14, 0)
 Call randomGenerate
Index1 = 2
     P = 0.09
     Beta = 1000 * moder
     NumberToGenerate = NumberEquals
 Call randomGenerate

Index1 = 3
     P = 0.09
     Beta = 1000 * moder
     NumberToGenerate = Excel.WorksheetFunction.Round(NumberEquals * 0.96, 0)
```

```
 Call randomGenerate

Index1 = 4
    P = 0.1
    Beta = 1100 * moder
    NumberToGenerate = Excel.WorksheetFunction.Round(NumberEquals * 1.06, 0)
 Call randomGenerate

'Index1 = 5
'    P = 0.08
'    Beta = 1100 * moder
'    NumberToGenerate = NumberEquals
' Call randomGenerate
'
'Index1 = 6
'    P = 0.08
'    Beta = 1100 * moder
'    NumberToGenerate = NumberEquals
' Call randomGenerate
'
'Index1 = 7
'    P = 0.08
'    Beta = 1100 * moder
'    NumberToGenerate = NumberEquals
' Call randomGenerate
'
'Index1 = 8
'    P = 0.08
'    Beta = 1100 * moder
'    NumberToGenerate = NumberEquals
' Call randomGenerate
'
'Index1 = 9
'    P = 0.13
'    Beta = 1400 * moder
'    NumberToGenerate = NumberEquals
' Call randomGenerate

End Sub

Sub randomGenerate()
   'random process
    Exposures = 0
    polcounts = 0
    Losses = 0
    LossesSquared = 0
For i = 1 To NumberToGenerate
If i < 0.43 * NumberToGenerate + 0.1 * Rnd() * NumberToGenerate Then
 b = Rnd()
    If b < P Then
    Frequency = 1
    Else: Frequency = 0
    End If

    a = Rnd()
    Loss = Excel.WorksheetFunction.GammaInv(a, Alpha, Beta)
    Loss = Frequency * Excel.WorksheetFunction.Round(Loss, 0)
    If Loss < Deductible Then
      Loss = 0
    End If
    If Loss > Limit Then
      Loss = Limit
    End If
    LossSquare = Loss ^ 2
    Losses = Losses + Loss
    LossesSquared = LossesSquared + LossSquare
    Exposures = Exposures + 1
    polcounts = polcounts + 1
Else
 b = Rnd()
    If b < P Then
    Frequency = 1
    Else: Frequency = 0
    End If

    a = Rnd()
    Loss = Excel.WorksheetFunction.GammaInv(a, Alpha, Beta)
    Loss = Frequency * Excel.WorksheetFunction.Round(Loss, 0)
    If Loss < Deductible Then
      Loss = 0
    End If
```

```
      If Loss > Limit Then
        Loss = Limit
      End If
      LossSquare = Loss ^ 2
      Losses = Losses + Loss
      LossesSquared = LossesSquared + LossSquare
      Exposures = Exposures + 1
      polcounts = polcounts + 1
   b = Rnd()
      If b < P Then
      Frequency = 1
      Else: Frequency = 0
      End If

      a = Rnd()
      Loss = Excel.WorksheetFunction.GammaInv(a, Alpha, Beta)
      Loss = Frequency * Excel.WorksheetFunction.Round(Loss, 0)
      If Loss < Deductible Then
        Loss = 0
      End If
      If Loss > Limit Then
        Loss = Limit
      End If
      LossSquare = Loss ^ 2
      Losses = Losses + Loss
      LossesSquared = LossesSquared + LossSquare
      Exposures = Exposures + 1
End If
Next
      SQL = "insert into data   (index1, polcounts, exposures, losses, LossesSquared)
values ("
      SQL = SQL & Index1 & "," & polcounts & "," & Exposures & "," & Losses & "," &
LossesSquared & ")"
      dber.Execute (SQL)
'start excel output -- this is so you can see the numbers generated.
countExcel = countExcel + 1
Range("a" & countExcel).Value = iter1
Range("b" & countExcel).Value = Index1
Range("c" & countExcel).Value = polcounts
Range("d" & countExcel).Value = Exposures
Range("e" & countExcel).Value = Losses
Range("f" & countExcel).Value = LossesSquared

Range("g" & countExcel).Value = "=RC[-1]/RC[-3]-(RC[-2]/RC[-3])^2"
Range("h" & countExcel).Value = "=IF(R[1]C[-6]>RC[-6],SQRT(RC[-1]/RC[-4]+R[1]C[-
1]/R[1]C[-4]),"""")"
Range("i" & countExcel).Value = "=RC[-4]/RC[-5]"
Range("j" & countExcel).Value = "=IF(R[1]C[-8]>RC[-8],(RC[-1]-R[1]C[-1])/RC[-2],"""")"
Range("k" & countExcel).Value = "=IF(R[1]C[-9]>RC[-9],NORMSDIST(RC[-1]),"""")"
Range("l" & countExcel).Value = "=IF(R[1]C[-10]>RC[-10],IF(RC[-1]>0.5,1-RC[-1],RC[-
1]),"""")"
Range("m" & countExcel).Value = "=IF(R[1]C[-11]>RC[-11],IF(RC[-
1]<0.05,""Significant"",""""),"""")"



'end excel output
    'end random process
End Sub
```

REFERENCES

[1] Richard G. Woll, "A Study of Risk Assessment," *Proceedings of the Casualty Actuarial Society*, Casualty Actuarial Society (1979) Vol. LXVI at Page 84 (Casualty Actuarial Society: Arlington, Virginia), See Http://www.casact.org/pubs/proceed/proceed79/index.htm

[2] R.J. Finger, "Risk Classification," *Foundations of Casualty Actuarial Science (Fourth Edition)* Pages 308-318 (Casualty Actuarial Society: Arlington, Virginia), See Http://www.casact.org/admissions/syllabus/2003/ch6.pdf

[3] Stuart A Klugman, Harry H. Panjer, Gordon E. Willmot, *Loss Models: From Data to Decisions* Pages 464-471; (Wiley-Interscience Publication, John Wiley & Sons, Inc. 1998)

[4] Romel G. Salam, "Reinventing Risk Classification--A Set Theory Approach," Ratemaking Call Papers, Casualty Actuarial Society (2002.) See Http://www.casact.org/pubs/forum/02wforum/02wftoc.htm

Incorporation of Fixed Expenses

Geoffrey Todd Werner, FCAS, MAAA

INCORPORATION OF FIXED EXPENSES

ABSTRACT

When setting rates, actuaries must include all of the costs of doing business, including underwriting expenses. Actuaries generally divide the underwriting expenses into two groups: fixed and variable. This paper addresses the incorporation of fixed expenses in the calculation of the actuarial indication. More specifically, the paper describes how the generally accepted method for including fixed expenses overstates or understates the actuarial indication. The materiality of the distortion depends on the magnitude of past rate changes, premium trend, and variations in average premiums for multi-state companies. For the example included, the generally accepted procedure overstated the indication by +1.8 percentage points. Finally, the paper suggests an alternative procedure that addresses the distortions.

ACKNOWLEDGMENT

While there were many individuals who helped me with the paper, I would like to specifically thank Catherine Taylor, Chris Norman, Heather McIntosh, Lisa Sukow, and Lisa Thompson for their efforts. Each of them provided significant assistance on the substance and/or style of this paper.

INCORPORATION OF FIXED EXPENSES

INTRODUCTION

The role of a pricing actuary is to set rates that provide for the expected future amount of all costs associated with the transfer of risk. Historically, actuarial literature has focused either on the larger costs of doing business (e.g., losses) or the more complex topics (e.g., profit provisions). Thus, there is relatively little literature dealing with the treatment of underwriting expenses.

Actuaries generally divide underwriting expenses into two groups: fixed and variable. Fixed expenses are those expenses that are assumed to be the same for each exposure, regardless of the size of the premium (i.e., the expense is a constant dollar amount for each risk). Typically, overhead costs associated with the home office are considered a fixed expense.[1]
Variable expenses are those expenses that vary directly with premium; in other words, the expense is a constant percentage of the premium. Premium taxes and commissions are two good examples of variable expenses.

---

[1] It is likely that some of these expenses do bear some relationship to risk and may vary at least slightly with premium. Activity-based cost studies may be able to verify the true relationship and appropriate adjustments can be made.

This paper discusses the often-overlooked portion of the premium, the fixed expenses. Specifically, the paper addresses:

- The generally accepted method for calculating a fixed expense provision and including it within the overall statewide rate level indication.
- Potential distortions that may make the current methodology misstate the actuarial indication.
- An alternative procedure for calculating and incorporating a fixed expense provision.

CURRENT METHOD

Calculation of Projected Fixed and Variable Expense Provision

A review of filings from several P&C personal lines insurers confirms that most actuaries use a method similar to the one outlined by David Schofield in "Going from a Pure Premium to a Rate" to calculate a fixed expense provision and expense fee [4]. Basically, the procedure assumes historical expense ratios (i.e., historical expenses divided by historical premiums) are the best estimate of projected expenses.

The first step of his procedure is to determine the percentage of premium attributable to expenses for each of the expense categories. To accomplish this, actuaries generally relate historical expenses to either written or earned premium for that same historical experience period. The choice of premium depends on whether the actuary believes the expenses are generally incurred at the onset of the policy or throughout the policy. Written premium is used in the former case and earned premium is used in the latter case. Once the appropriate ratios are determined for each type of expense, the ratios are then split into a fixed expense ratio and a variable expense ratio based on all available expense data, regulations, and judgment.

Exhibit 1 shows the relevant exhibits for such a procedure. The data used is Homeowners data adjusted so that the three-year historical expense ratios (i.e., expenses divided by premiums) are approximately equal to the three-year industry historical expense ratios.

Exhibit 1-A displays three years of historical expense ratios. All of this information can be derived from the applicable Insurance Expense Exhibits (IEE's) and Statutory Page 14's, if necessary. The IEE's and Statutory Page 14's may not be in the finest level of detail desired. For example, the Homeowners data includes Renters data and Mobile Homes data. Ideally, the actuary can access and use the source expense data to get the data corresponding to the product being priced. Of course, the actuary should always balance the additional cost of obtaining such data against the additional accuracy gained. In this case, the company assumes all expenses, except General Expenses, are incurred at the onset of the policy and divides them by written premium. General Expenses are assumed to be incurred throughout the policy period, and thus are divided by earned premium.

Typically, the data used (i.e., countrywide or state) also varies by type of expense. Other Acquisition, and General Expenses are generally assumed to be uniform across all locations, and hence can be handled using countrywide figures that can be found in the IEE. The handling of Commissions and Brokerage varies from carrier to carrier with some carriers using state specific data and some using countrywide figures; the treatment should be based on the variation of the company's commission plans by location. Taxes, Licenses and Fees vary by state; therefore, they are typically based on state data from the applicable Statutory Page 14. Ideally, the company can break the category into taxes, which is a variable expense, and licenses and fees, which are typically treated as fixed expenses.[2]

The following chart summarizes the information:

Expense	Data Used	Divided By
General Expense	Countrywide	Earned Premium
Other Acquisition	Countrywide	Written Premium
Commissions and Brokerage	Countrywide/State	Written Premium
Taxes	State	Written Premium
Licenses & Fees	State	Written Premium

Once the historical ratios are calculated, the actuary chooses a selected provision for each expense type. Generally, the selection is based on either the latest year or a multi-year average; however, there are several things that may affect the selection:

- If the actuary is aware of a future change in the expenses, the new figure should be used. For example, if the commission structure is changing, the actuary should use the expected commission percentage, not the historical percentage.
- If there was a one-time shift in expense levels during the experience period, the expected future expense level should be used. For example, if productivity gains led to a significant

[2] Licenses and Fees tend to be a smaller portion of the overall Taxes, Licenses, and Fees category. Thus, if a

reduction in necessary staffing levels during the historical experience period, then the selected ratios should be based on the ratios after the reduction rather than the all-year average.

- If there were non-recurring expense items during the historical period, the actuary should examine the materiality and nature of the expense to determine how to best incorporate the expense in the rates—if at all. If the aggregate dollars spent are consistent with dollars spent on similar non-recurring projects in other years, the expense ratios should be similar and no adjustment is warranted. If, however, the expense item represents an extraordinary expense, then the actuary must decide to what extent it should be included. Assume, for example, the extraordinary expense is from a major systems project to improve the policy issuance process. That project clearly benefits future policyholders and should be included in the rates. Assuming the new system will be used for a significant length of time, it may be appropriate to dampen the impact of the item and spread the expense over a period of several years. If the actuary consistently selects the three-year average, the expense will automatically be spread over three years assuming rates are revised annually.[3] On the other hand, the non-recurring expense may be caused by something for which the actuary determines it is inappropriate to charge future policyholders. If so, the actuary should exclude the expense from the ratemaking data altogether. In that case, the expense is basically funded by existing surplus.

---

company is unable to split them out, the inclusion of these with variable expenses will not have a material effect.
[3] This assumes all of the expense is booked in that year. Statutory accounting guidelines allow some expenses to be amortized over several years. If the extraordinary expense is amortized over three years, then the use of a three-year average will actually spread the expense over five years. The three-year average expense ratio will increase for the first three years and decrease for the last two years.

- Finally, a few states place restrictions on which expenses can be included for the purpose of determining rates. For example, Texas does not allow insurers to include charitable contributions or lobbying expenses. These expenses must be excluded from the calculation of the historical expense ratios when performing the analysis for that state. If such expenses are recurring, overall future income will be reduced by that state's proportion of the expenses.

In the example on Exhibit I-A, the data is fairly stable and there were no extraordinary expenses; therefore, a three-year straight average is selected.

Once the expense ratios are selected, they are divided into fixed and variable ratios. Ideally, the company has detailed expense data and can do this directly or has activity-based cost studies that help split the expenses appropriately. In the absence of any such data, the actuary should consult with other insurance professionals within the company to arrive at the best possible assumptions given the company's allocation of expenses. In this example, the company assumes 75% of the General Expenses and Other Acquisition costs and 100% of the Licenses and Fees are fixed. All other expenses are assumed to be variable. Some sensitivity testing was performed on these selections. For the example included, the difference in the indications between assuming the aforementioned percentages of the General Expenses, Other Acquisition, and Licenses and Fees were fixed and assuming 100% of those expenses were fixed is not material. The exact impact will vary and depend on the magnitude of the expense ratios.

The fixed expense ratio represents the fixed expenses for the historical time period divided by premium written or earned during that same time period. Often, companies trend this ratio to account for expected growth in fixed expenses. Some companies use internal expense data to select an appropriate trend. Given the volatility of internal data, many companies use government indices (e.g., Consumer Price Index, Employment Cost Index, etc.) and knowledge of anticipated changes in internal company practices to estimate an appropriate trend. Exhibit 1-B displays one such methodology. Basically, the indicated trend is a weighting of the Employment Cost Index and the Consumer Price Index. The weights are based on the percentage that salaries represent of the total expenditures for the two largest fixed expense categories, Other Acquisition and General Expenses. These weights can be determined directly from data contained in the IEE. The selected fixed expense ratio will be trended from the average date expenses were incurred in the historical expense period to the average date expenses will be incurred in the period the rates are assumed to be in effect (see Appendix A).[4] After making that adjustment, the ratio is often called the projected (or trended) fixed expense provision.

Variable expenses and profit are a constant percentage of the premium. That selected percentage will apply to the premiums from policies written during the time the rates will be in effect. Thus, there is no need to trend that ratio. The ratio is called the variable expense provision.

[4] When multi-year historical ratios are used, there is often no trending to bring each year's ratio to the same expense and premium levels before making a selection. If the expenses and average premiums are changing at the same rate, then the two would offset each other and the ratios would remain constant. However, if the expense trend exceeds the change in average premium (or the change in average premium exceeds the expense trend), this would tend to result in increasing (decreasing) ratios over the historical period. The materiality of this distortion depends on the magnitude of the difference between the trends.

Calculation of Statewide Indicated Change

Exhibit 1-C shows the most commonly found method of incorporating the fixed expense provision within the calculation of the indicated statewide rate level change. The general formula for the statewide (SW) indicated change based on the loss ratio method is as follows:

$$\text{SW Indicated Change} = \frac{\text{Projected Loss Ratio + Projected Fixed Expense Provision}}{\text{1.00-Variable Expense Provision-Profit \& Contingency Provision}} - 1.00$$

The projected fixed expense provision and the variable expense provision are calculated as discussed in the prior section. Much literature is dedicated to the determination of loss ratios and profit and contingency provisions; thus, they will not be discussed further here.

POTENTIAL DISTORTIONS

There are a few items that can cause the preceding methodology to create inaccurate and inequitable indicated rate changes.

First, rate changes[5] can impact the historical expense ratios and lead to an excessive or inadequate overall rate indication. The historical fixed expense ratios are based on written and

[5] The term "rate changes" (or premium level changes) is intended to refer to changes resulting from an increase or decrease in the premiums. The term is not intended to be used interchangeably with "rate level changes" which can be caused by premium changes, coverage changes, or both. If a rate level change is caused solely by a change in coverage, it may or may not impact the appropriateness of the historical expense ratios. If the actuary adjusts the losses to account for coverage level changes, there will not be a distortion. If, however, the actuary adjusts premiums to account for such changes, the distortion will still exist.

earned premium during that time period. To the extent that there are rate increases (or decreases) that only impact a portion of the premium in the historical time period or were implemented after the historical period entirely, the current procedure will tend to overstate (or understate) the expected fixed expenses. The materiality of the distortion depends on the magnitude of rate changes not fully reflected in the historical countrywide premiums. Also, utilizing three-year historical expense ratios increases historical experience period thereby increasing the chances of rate changes not being fully reflected in the historical premiums. One potential solution for the distortion caused by rate changes is to restate the historical written or earned premiums at current rates.

Second, significant premium trend between the historical experience period and the projected period can lead to an excessive or inadequate overall rate indication.[6] Again, the historical expenses are divided by the written and earned premium during that time period. To the extent that there have been distributional shifts that have increased the average premium (e.g., higher amounts of insurance) or decreased the average premium (e.g., higher deductibles), this methodology will tend to overstate or understate the estimated fixed expenses, respectively. The magnitude of overstatement or understatement depends on the magnitude of the premium trend. Utilizing three-year historical expense ratios increases the impact of premium trend by increasing the amount of time between the historical and projected periods. A potential solution for this is to trend the historical premiums to prospective levels.

[6] This assumes the premium trend is due to changes that do not proportionately increase (or decrease) the fixed expenses. While this is the most common scenario, there may be situations that deviate from this assumption. For example, assume a company is pursuing an insurance-to-value (ITV) effort with an external inspection company. Presumably, the additional expenses incurred will lead to better ITV and higher average premiums. Thus, both average premiums and average expenses would be increasing. In a case like this, the actuary should determine the impact, decide if this is a one-time shift, and adjust the selections accordingly.

Third, this methodology can create inequitable rates for regional or nationwide carriers because it uses countrywide expense ratios[7] and applies them to state projected premiums to determine the expected fixed expenses. In other words, fixed expenses are allocated to each state based on premium. The average premium level in states can vary due to overall loss cost differences (e.g., coastal states tend to have higher overall homeowners loss costs) as well as distributional differences (e.g., some states have a significantly higher average amounts of insurance than other states). If there exists significant variation in average rates across the states, a disproportionate share of projected fixed expenses will be allocated to the higher-than-average premium states. Thus, the estimated fixed expenses will be overstated in higher-than-average premium states and understated in the lower-than-average average premium states. If a company tracks fixed expenses by state and calculates fixed expense ratios for each state, then this distortion will not exist.

PROPOSED METHODOLOGY

By assumption, fixed expenses are assumed to be constant for each exposure and are not assumed to vary with premium. The proposed methodology uses the concepts outlined by Diana Childs and Ross Currie in "Expense Allocation in Insurance Ratemaking" [1]. In essence, historical fixed expenses are divided by historical exposures rather than premium. Exhibit 2

[7] State-specific data is usually used for taxes, licenses, and fees. However, these expenses are relatively small compared to the other expenses that are generally evaluated on a countrywide level.

displays this procedure.

Calculation of Projected Fixed and Variable Expense Provisions

Exhibit 2-A, Sheet i shows the development of the fixed and variable expenses for the General Expense category. The total expenses for the category can be taken directly from the IEE. The total expenses are split into variable and fixed expenses. Ideally, the expenses are maintained at a level of detail that allows an accurate allocation between the variable and fixed expense categories. Typically, the total expenses are split using percentages based on internal company data and/or actuarial judgment. This example uses the same percentages assumed in the current procedure (75% of General Expenses and Other Acquisition costs and 100% of Licenses and Fees are fixed and all other expenses are variable).[8]

The total fixed expenses are then divided by the exposures[9] for that same time period. As General Expenses are assumed to be incurred throughout the policy, the expense dollars are divided by earned exposures, rather than written, to determine an average expense per exposure for the indicated historical period. The average expense figures are trended using the same approach discussed earlier in the paper (see Exhibit 1-B). All of the average expense amounts are trended from the average date they were incurred in the historical period to the average date

[8] Note, if premiums and expenses are changing at different rates, then the percentage that fixed expenses are of total expenses will change over time, but that does not result in a material distortion. See Appendix B for more discussion on this issue.

[9] House-years were used as the exposure unit for the example in the paper. Using amount-of-insurance years as an exposure base will lead to distortions similar to those caused by the current procedure, if there are significant differences in amounts of insurance over time and among various locations.

expenses will be incurred in the period the rates will be in effect.[10] Once the projected expenses per exposures are determined, the actuary then must select an appropriate figure.

As with the current procedure, the selection will generally be based on either the latest year or a multi-year average. Consistent values for the projected average expense per exposure imply expenses are increasing or decreasing proportionately to exposures. This makes intuitive sense for many expense categories (e.g., full-time employee costs), but may not be accurate for all fixed expenses due to economies of scale. If the company is growing and the projected average expense per exposure is declining steadily each year, then it is an indication that the selected expense trend may be too high and/or that expenses may not be increasing as quickly as exposures due to economies of scale. If the decline is significant and the actuary believes it is because of economies of scale, then the selection should be adjusted to include the impact of economies of scale given expected growth in the book.[11] As mentioned earlier, non-recurring expense items, one-time changes in expense levels, or anticipated changes in expenses should be considered in making the selection. In the example shown the figures are stable and the three-year average is selected to facilitate comparisons with the results of the current procedure.

Exhibit 2-A, Sheets i-iv show the calculations for each of the major expense categories. The following chart summarizes the data used:

[10] In the example, the same trend period is used for all expense categories to maintain consistency with the current procedure. See Appendix A for more discussion on this issue.

[11] If the selected expense trend is based on historical internal expense data (e.g., historical changes in average expense per exposure) rather than external indices, then the trend would implicitly include the impact of economies of scale in the past. Assuming the impact of economies of scale will be the same as in the past, the projected average expense per exposure should be consistent and no further adjustment would be necessary.

		Divided By	
Expense	**Data Used**	**Fixed**	**Variable**
General Expense	Countrywide	Earned Exposures	Earned Premium
Other Acquisition	Countrywide	Written Exposures	Written Premium
Commissions and Brokerage	Countrywide/State	--	Written Premium
Taxes	State	--	Written Premium
Licenses & Fees	State	Written Exposures	--

Exhibit 2-B summarizes the results of the analysis of the fixed and variable portions of each major expense group.

Calculation of Statewide Indicated Change

The most straightforward way to calculate the indicated change is displayed on Exhibit 2-C. The statewide required projected average premium is calculated as follows:

$$\text{SW Projected Average Required Premium} = \frac{\text{SW Projected Average Loss \& LAE Per Exposure} + \text{Projected Average Fixed Expense Per Exposure}}{\text{1.00-Variable Expense Provision-Profit \& Contingency Provision}}$$

That figure is compared to the statewide projected average premium at present rates to determine the statewide indicated change:

$$\text{SW Indicated Change} = \frac{\text{SW Projected Average Required Premium}}{\text{SW Projected Average Premium at Present Rates}} - 1.00$$

Alternatively, the projected average fixed expense per exposure can be converted to a projected fixed expense provision by dividing the projected average fixed expense per exposure by the statewide projected average premium at present rates. That figure can then be used within the same loss ratio indication formula provided earlier:

$$\text{SW Indicated Change} = \frac{\text{Projected Loss Ratio + Projected Fixed Expense Provision}}{\text{1.00-Variable Expense Provision-Profit \& Contingency Provision}} - 1.00$$

Calculation of Expense Fees

Some insurers may have expenses fees or minimum premiums. If that is the case, this procedure directly lends itself to the determination of such values.

Exhibit 2-D displays the necessary calculations for an expense fee. The projected average fixed expense per exposure has already been calculated. To calculate an expense fee, that figure needs to be increased to cover the variable items (variable expenses and profit) associated with the fixed portion of the premium. This is accomplished simply by dividing the figure by the variable permissible loss ratio (i.e., 1.00-variable expense provision-profit provision).

To determine a minimum premium, the amount necessary to cover the expenses should be combined with a minimum provision for losses.

CURRENT METHODOLOGY VERSUS PROPOSED METHODOLOGY

This section algebraically shows the difference in the projected fixed expense dollars calculated under the two different methodologies. The formula for calculating the total dollars of projected statewide fixed expenses using the current methodology is as follows[12]:

$$\text{Proj SW Fixed Expenses}_{Curr} = \frac{\text{Historical CW Fixed Expenses}}{\text{Historical CW Premium}} * \text{Expense Trend Factor} * \text{Proj SW Premium}$$

The formula for calculating the projected statewide fixed expenses collected using the proposed methodology is as follows:

$$\text{Proj SW Fixed Expenses}_{Prop} = \frac{\text{Historical CW Fixed Expenses}}{\text{Historical CW Exposures}} * \text{Expense Trend Factor} * \text{Proj SW Exposures}$$

Dividing the first formula by the second highlights the relative difference between the fixed expenses produced by the two procedures:

$$\frac{\text{Proj SW Fixed Expenses}_{Curr}}{\text{Proj SW Fixed Expenses}_{Prop}} = \frac{\text{Historical CW Exposures}}{\text{Historical CW Premium}} * \frac{\text{Proj SW Premium}}{\text{Proj SW Exposures}}$$

Equivalently,

$$\frac{\text{Proj SW Fixed Expenses}_{Curr}}{\text{Proj SW Fixed Expenses}_{Prop}} = \frac{\text{Proj SW Avg Premium}}{\text{Historical CW Avg Premium}}$$

[12] The following section only deals with the categories of expenses that use the countrywide expenses. Taxes, Licenses, and Fees are not addressed. Those expenses represent a relatively small portion of the total expense dollars.

Multiplying by unity (i.e., Proj CW Avg Premium/Proj CW Avg Premium),

$$\frac{\text{Proj SW Fixed Expenses}_{\text{Curr}}}{\text{Proj SW Fixed Expenses}_{\text{Prop}}} = \frac{\text{Proj SW Avg Premium}}{\text{Historical CW Avg Premium}} * \frac{\text{Proj CW Avg Premium}}{\text{Proj CW Avg Premium}}$$

Rearranging the terms,

$$\frac{\text{Proj SW Fixed Expenses}_{\text{Curr}}}{\text{Proj SW Fixed Expenses}_{\text{Prop}}} = \frac{\text{Proj CW Avg Premium}}{\text{Historical CW Avg Premium}} * \frac{\text{Proj SW Avg Premium}}{\text{Proj CW Avg Premium}}$$

Since

$$\text{Proj CW Avg Premium} = \text{Historical CW Avg Premium} * \text{Premium Trend Factor} * \text{On-level Factor}$$

We have

$$\frac{\text{Proj SW Fixed Expenses}_{\text{Curr}}}{\text{Proj SW Fixed Expenses}_{\text{Prop}}} = \text{Premium Trend Factor} * \text{On-level Factor} * \frac{\text{Proj SW Avg Premium}}{\text{Proj CW Avg Premium}}$$

The difference between the fixed expenses produced by the two methodologies is driven by premium trend, on-level factors, and the relationship of the statewide average premium to the countrywide average premium. These represent the three distortions mentioned earlier. Thus, the proposed methodology is not affected by these three distortions.

Exhibit 3 shows the impact on the overall indication by location for the two methodologies (Exhibit 3-A lists the information in table form and Exhibit 3-B represents the data graphically). This information is included to show two items: the total amount the current procedure overstates/(understates) the overall indication relative to the proposed procedure and the variation of the overstatement/(understatement) by location. The former tells us about the impact on the accuracy of the overall countrywide indication, while the latter is more indicative of equity issues among states.

An examination of the "Countrywide" line on Exhibit 3-A shows the current procedure overstates the premium needed to cover projected fixed expenses by +1.8 percentage points relative to the proposed procedure. During the historical period used, homeowners insurance rates were being increased and the overall premium trend was slightly positive. For these two reasons, the proposed procedure determines a fixed expense provision that is less than that produced by the current procedure.

A survey of the impact by location shows significant variation (a high of +10.6 percentage points to a low of -8.3 percentage points). The location specific differences are driven by the differences in average projected premiums at present rates (PPR). The average projected PPR can and does vary significantly from location to location due to the overall cost of doing business in the states as well as differing distributions of high and low risk insureds in the states. The relationship of each state's average projected PPR to the countrywide average projected PPR is included. In general, the higher the average projected PPR, the more the current procedure overstates the indication relative to the proposed procedure.

As mentioned earlier, the expense ratios in the example approximate the Homeowners industry three-year expense ratios. To the extent that an individual company has a greater (or lesser) percentage of fixed expenses than the industry average, then the impacts will be larger (or smaller). Additionally, the results depend on the rate changes, premium trends, and statewide rate relativities underlying the data.

OTHER CONSIDERATIONS/FUTURE ENHANCEMENTS

While the procedure does correct for the three distortions mentioned, there are still some concerns that are not addressed.

First, the proposed procedure—like the current procedure—requires the actuary to split the expenses into fixed and variable categories. Today, this is generally done judgmentally. Perhaps future activity-based cost studies will more accurately segregate expenses. As mentioned earlier, sensitivity testing revealed the overall indication is not materially impacted by moderate swings in the categorization of expenses.

Second, the proposed procedure essentially allocates countrywide fixed expenses to each state based on the by-state exposure distribution (as it assumes fixed expenses do not vary by exposure). In reality, average fixed expense levels may vary by location (e.g., advertising costs may be higher in some locations than others). If a regional or nationwide carrier feels the variation is material, the company should collect data at a finer level and make the appropriate adjustments. Once again, the cost of the data collection should be balanced against the additional accuracy gained.

Third, some expenses considered fixed probably vary slightly with premium. For example, policies for coastal homes may be more costly to service than other homes. Further studies may uncover a more accurate quantification of this relationship. However, assuming the expenses are "nearly" fixed, the resulting inequity is not material.

Fourth, some expenses considered fixed vary by other characteristics. For example, fixed expenses may vary between new and renewal business. This only affects the overall statewide indication if the distribution of risks for that characteristic is changing dramatically and/or varies significantly by state. Even if there is no impact on the overall indication, any material fixed expense cost difference not reflected in the rates will impact the equity of the two groups. To make rates equitable for the example of new versus renewal business, material differences in new and renewal provisions should be reflected with consideration given to varying persistency levels as described by Sholom Feldblum in "Asset Share Pricing for Property and Casualty Insurers" [3].

Finally, the existence of economies of scale in a changing book will lead to increasing or decreasing projected average expense per exposure figures. Further studies may reveal techniques for better approximating the relationship between changes in exposures and expenses and capturing the impact of economies of scale. Until then, internal expense trend data and actuarial judgment should suffice for incorporating the impact of economies of scale.

CONCLUSION

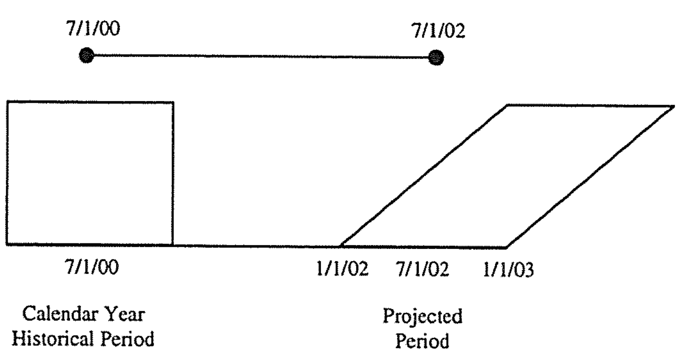
The prevailing methodology for incorporating fixed expenses in the statewide indication has some methodological flaws. Those flaws can lead to overstated or understated actuarial indications. While this paper describes a simple alternative that corrects the three weaknesses mentioned, there are still improvements that can be made.

APPENDIX A

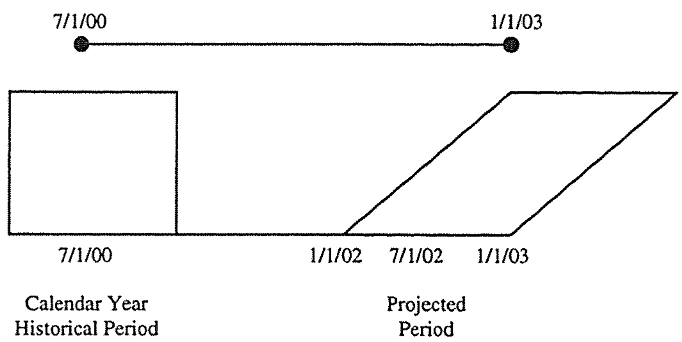
Trending Periods

Expenses should be trended from the average date they were incurred in the historical period to the average date they will be incurred in the projected period. Actuaries generally make the simplifying assumption that expenses are either incurred at the inception of the policy or are incurred evenly throughout the policy period. When using calendar year historical expense data, the trend periods should be different for the two different types of expenses.

First, expenses that are incurred at the inception of the policy should be trended from the average written date in the historical period to the average written date in the projection period. The following figure shows the resulting trend period assuming annual policies, a steady book of business, and that the projected rates will be in effect for one year:



Second, expenses that are incurred evenly throughout the policy period should be trended from the average earned date in the historical period to the average earned date in the projection period. The following figure shows the resulting trend period assuming annual policies, a steady book of business, and that the projected rates will be in effect for one year:



As can be seen by the figures, under our assumptions, expenses incurred throughout the policy are trended 6 months longer than expenses incurred at inception. Indications do not generally include different trend periods for the different expenses. Presumably, a common trend period is used for simplicity, as this distinction does not have a material impact. The exhibits in the paper use a common trend period.

APPENDIX B

Does the Percentage that Fixed Expense Represent of Total Expenses Vary Over Time?

In both the current and proposed procedure, the actuary must separate the expenses into fixed and variable expenses. Since detailed expense data may not be available, the actuary may have to use a judgmentally selected percentage to split the expenses from the Annual Statement.

Generally, that same percentage is applied to the expenses for each of the years in the historical period. If the change in average premium does not equal the fixed expense trend, then fixed and variable expenses will be growing at different rates. Thus, the percentages that fixed expenses and variable expenses represent of total expenses will change over time.

Some sensitivity analysis was performed to determine the impact on the indications of a change in the percents that fixed and variable expenses are of total expenses. For the sensitivity analysis, the same example was used with the assumption the percentage was accurately determined in year 1. Even with the very unlikely assumption that average premiums were changing at a rate in excess of +10 percentage points differently than expenses, the indications were only impacted by about +0.2 percentage points. In reality, premiums and expenses would likely be changing at a more equivalent rate. So, using a constant percentage for the three-year period is a reasonable assumption.

REFERENCES

[1] Childs, Dianna and Currie, Ross. "Expense Allocation in Insurance Ratemaking." CAS Discussion Paper Program 1980 Volume: May.

[2] Committee on Ratemaking Principles, Statement of Principles Regarding Property and Casualty Insurance Ratemaking, Casualty Actuarial Society.

[3] Feldblum, Sholom. "Asset Share Pricing for Property and Casualty Insurance." CAS Forum 1993 Volume: Special Edition.

[4] Schofield, David. "Going from a Pure Premium to a Rate." CAS Forum 2000 Volume: Winter.

Exhibit 1-A

Company
State XX Homeowners
Projected Fixed and Variable Expense Provisions
Current Method

		Year1	Year2	Year3	3-Year Average	Selected
(1) General Expenses						
a CW Expenses		$ 26,531,974	$ 28,702,771	$ 31,195,169		
b CW Earned Premium		$ 450,000,000	$ 490,950,000	$ 545,250,000		
c Ratio [(a)/(b)]		5.9%	5.8%	5.7%	5.8%	5.8%
d % Assumed Fixed						75.0%
e Fixed Expense % [(c)x(d)]						4.4%
f Variable Expense % [(c)x(1.0-(d))]						1.5%
(2) Other Acquisition						
a CW Expenses		$ 41,758,296	$ 45,612,462	$ 49,582,543		
b CW Written Premium		$ 468,850,000	$ 515,550,000	$ 577,900,000		
c Ratio [(a)/(b)]		8.9%	8.8%	8.6%	8.8%	8.8%
d % Assumed Fixed						75.0%
e Fixed Expense % [(c)x(d)]						6.6%
f Variable Expense % [(c)x(1.0-(d))]						2.2%
(3) Licenses and Fees						
a State Expenses*		$ 1,157,006	$ 1,210,200	$ 1,321,419		
b State Written Premium*		$ 468,850,000	$ 515,550,000	$ 577,900,000		
c Ratio [(a)/(b)]		0.2%	0.2%	0.2%	0.2%	0.2%
d % Assumed Fixed						100.0%
e Fixed Expense % [(c)x(d)]						0.2%
f Variable Expense % [(c)x(1.0-(d))]						0.0%
(4) Commission and Brokerage						
a CW Expenses		$ 63,507,320	$ 69,832,993	$ 78,278,512		
b CW Written Premium		$ 468,850,000	$ 515,550,000	$ 577,900,000		
c Ratio [(a)/(b)]		13.5%	13.5%	13.5%	13.5%	13.5%
d % Assumed Fixed						0.0%
e Fixed Expense % [(c)x(d)]						0.0%
f Variable Expense % [(c)x(1.0-(d))]						13.5%
(5) Taxes						
a State Expenses*		$ 10,607,226	$ 9,917,093	$ 11,580,187		
b State Written Premium*		$ 468,850,000	$ 515,550,000	$ 577,900,000		
c Ratio [(a)/(b)]		2.3%	1.9%	2.0%	2.1%	2.1%
d % Assumed Fixed						0.0%
e Fixed Expense % [(c)x(d)]						0.0%
f Variable Expense % [(c)x(1.0-(d))]						2.1%
(6) Subtotal Fixed Expenses	[(1e)+(2e)+(3e)+(4e)+(5e)]					11.2%
(7) Fixed Expense Trend	[Exhibit 1, Sheet B]					3.4%
(8) Trend Period						3.00
(9) Fixed Expense Trend Factor	[(1.00+(7))^(8)]					1.1055
(10) Projected Fixed Expense Provision	**[(6)x(9)]**					**12.4%**
(11) Variable Expense Provision	**[(1f)+(2f)+(3f)+(4f)+(5f)]**					**19.3%**

*CW data used for example.

Exhibit 1-B

Company
Countrywide Homeowners
Calculation of Annual Expense Trend

(1) Employment Cost Index - Finance, Insurance & Real Estate,excluding Sales Occupations - (annual change over latest 2 years) U.S. Department of Labor	4.8%
(2) % of Other Acquisition and General Expenses used for Salaries and Employee Relations & Welfare - Insurance Expense Exhibit, Year3	50.0%
(3) Consumer Price Index, All Items - (annual change over latest 2 years)	1.9%
(4) Annual Expense Trend - { (1) * (2) } + { (3) * [100% - (2)] }	3.4%
Selected Annual Expense Trend	**3.4%**

Exhibit 1-C

Company
State XX Homeowners
Calculation of Indicated Rate Change
Current Method

(1) Projected Loss & LAE Ratio	64.7%
(2) Projected Fixed Expense Provision	12.4%
(3) Variable Expense Provision	19.3%
(4) Profit and Contingencies Provision	5.0%
(5) Variable Permissible Loss Ratio [100%-(3)-(4)]	75.7%
(6) Indicated Rate Change [(1)+(2)]/(5)-100%	1.8%

Exhibit 2-A
Sheet i

Company
State XX Homeowners
General Expenses
Proposed Method

	Year1	Year2	Year3	3-Year Straight Average	Selected
(1) Total CW General Expenses (IEE)	$ 26,531,974	$ 28,702,771	$ 31,195,169		
FIXED					
(2) Fixed General Expense as % of Total General Expense	75.0%	75.0%	75.0%		
(3) Fixed General Expense $	$ 19,898,981	$ 21,527,078	$ 23,396,377		
(4) Total CW Earned Exposures	625,500	666,500	696,000		
(5) Average Fixed General Expense Per Exposure	$ 31.81	$ 32.30	$ 33.62		
(6) Expense Trend	3.4%	3.4%	3.4%		
(7) Trend Period	4.00	3.00	2.00		
(8) Expense Trend Factor	1.1431	1.1055	1.0692		
(9) Projected Average Fixed General Expense Per Exposure	$ 36.36	$ 35.71	$ 35.95	$ 36.01	$ 36.01
VARIABLE					
(10) Variable General Expense as % of Total General Expense	25.0%	25.0%	25.0%		
(11) Variable General Expense $	$ 6,632,994	$ 7,175,693	$ 7,798,792		
(12) CW Earned Premium	$ 450,000,000	$ 490,950,000	$ 545,250,000		
(13) Variable General Expense %	1.5%	1.5%	1.4%	1.5%	1.5%

Notes:
(3) =(1)x(2)
(5) =(3)/(4)
(6) Exhibit 1-B
(7) 6/30/XX to midpoint of anticipated coverage period
(8) =[1.00+(6)]^(7)
(9) =(5)x(8)
(10) =100%-(2)
(11) =(1)x(10)
(13) =(11)/(12)

Exhibit 2-A
Sheet i

Company
State XX Homeowners
Other Acquisition
Proposed Method

	Year1	Year2	Year3	3-Year Straight Average	Selected
(1) Total CW Other Acq. Expenses (IEE)	$ 41,758,296	$ 45,612,462	$ 49,582,543		
FIXED					
(2) Fixed Other Acq. Expense as % of Total Other Acq. Expense	75.0%	75.0%	75.0%		
(3) Fixed Other Acq. Expense $	$ 31,318,722	$ 34,209,347	$ 37,186,907		
(4) Total CW Written Exposures	646,500	687,000	717,500		
(5) Average Fixed Other Acq. Expense Per Exposure	$ 48.44	$ 49.80	$ 51.83		
(6) Expense Trend	3.4%	3.4%	3.4%		
(7) Trend Period	4.00	3.00	2.00		
(8) Expense Trend Factor	1.1431	1.1055	1.0692		
(9) Projected Average Fixed Other Acq. Expense Per Exposure	$ 55.37	$ 55.05	$ 55.42	$ 55.28	$ 55.28
VARIABLE					
(10) Variable Other Acq. Expense as % of Total Other Acq. Expense	25.0%	25.0%	25.0%		
(11) Variable Other Acq. Expense $	$ 10,439,574	$ 11,403,116	$ 12,395,636		
(12) CW Written Premium	$ 468,850,000	$ 515,550,000	$ 577,900,000		
(13) Variable Other Acq. Expense %	2.2%	2.2%	2.1%	2.2%	2.2%

Notes:
(3) =(1)x(2)
(5) =(3)/(4)
(6) Exhibit 1-B
(7) 6/30/XX to midpoint of anticipated coverage period
(8) =[1.00+(6)]^(7)
(9) =(5)x(8)
(10) =100%-(2)
(11) =(1)x(10)
(13) =(11)/(12)

Exhibit 2-A
Sheet i

Company
State XX Homeowners
Taxes, Licenses & Fees
Proposed Method

	Year1	Year2	Year3	3-Year Straight Average	Selected
(1) Total State Taxes, Licenses & Fees (AS, Page 15)*	$ 11,764,232	$ 11,127,293	$ 12,901,606		
FIXED					
(2) Fixed Licenses & Fees Expense $	$ 1,157,006	$ 1,210,200	$ 1,321,419		
(3) Total State Written Exposures*	646,750	687,000	717,650		
(4) Average Fixed Licenses & Fees Expense Per Exposure	$ 1.79	$ 1.76	$ 1.84		
(5) Expense Trend	3.4%	3.4%	3.4%		
(6) Trend Period	4.00	3.00	2.00		
(7) Expense Trend Factor	1.1431	1.1055	1.0692		
(8) Projected Average Fixed Licenses & Fees Tax Expense Per Exposure	$ 2.05	$ 1.95	$ 1.97	$ 1.99	$ 1.99
VARIABLE					
(9) Variable Premium Tax Expense $	$ 10,607,226	$ 9,917,093	$ 11,580,187		
(10) State Written Premium*	$ 468,850,000	$ 515,550,000	$ 577,900,000		
(11) Variable Premium Tax Expense %	2.3%	1.9%	2.0%	2.1%	2.1%

Notes:
(4) =(2)/(3)
(5) Exhibit 1-B
(6) 6/30/XX to midpoint of anticipated coverage period
(7) =[1.00+(5)]^(6)
(8) =(4)x(7)
(11) =(9)/(10)

Company
State XX Homeowners
Commissions & Brokerage
Proposed Method

	Year1	Year2	Year3	3-Year Straight Average	Selected
(1) Total CW Comm. & Brok. Expenses (IEE)	$ 63,507,320	$ 69,832,993	$ 78,278,512		
FIXED					
(2) Fixed Comm. & Brok. Expense as % of Total Comm. & Brok. Expense	0.0%	0.0%	0.0%		
(3) Fixed Comm. & Brok. Expense $	$ -	$ -	$ -		
(4) Total CW Written Exposures	646,500	687,000	717,500		
(5) Average Fixed Comm. & Brok. Expense Per Exposure	$ -	$ -	$ -		
(6) Expense Trend	3.4%	3.4%	3.4%		
(7) Trend Period	4.00	3.00	2.00		
(8) Expense Trend Factor	1.1431	1.1055	1.0692		
(9) Projected Average Fixed Comm. & Brok. Expense Per Exposure	$ -	$ -	$ -	$ -	$ -
VARIABLE					
(10) Variable Comm. & Brok. Expense as % of Total Comm. & Brok. Expens	100.0%	100.0%	100.0%		
(11) Variable Comm. & Brok. Expense $	$ 63,507,320	$ 69,832,993	$ 78,278,512		
(12) CW Written Premium	$ 468,850,000	$ 515,550,000	$ 577,900,000		
(13) Variable Comm. & Brok. Expense %	13.5%	13.5%	13.5%	13.5%	13.5%

Notes:
(3) =(1)x(2)
(5) =(3)/(4)
(6) Exhibit 1-B
(7) 6/30/XX to midpoint of anticipated coverage period
(8) =[1.00+(6)]^(7)
(9) =(5)x(8)
(10) =100%-(2)
(11) =(1)x(10)
(13) =(11)/(12)

Exhibit 2-B

Company
State XX Homeowners
Projected Fixed and Variable Expense Provisions
Proposed Method

	Fixed	Variable
(1) General Expenses	$ 36.01	1.5%
(2) Other Acquisition	$ 55.28	2.2%
(3) Taxes, Licenses & Fees	$ 1.99	2.1%
(4) Commissions & Brokerage	$ -	13.5%
(5) Total	$ 93.28	19.3%

Company
State XX Homeowners
Calculation of Indicated Rate Change
Proposed Method

(1) Statewide Projected Average Premium at Present Rates*	$ 850.59
(2) Statewide Projected Loss & LAE Ratio	64.7%
(3) Statewide Projected Average Loss & LAE (1)x(2)	$ 550.33
(4) Projected Average Fixed Expense Per Exposure	$ 93.28
(5) Variable Expense Provision	19.3%
(6) Profit and Contingencies Provision	5.0%
(7) Variable Permissible Loss Ratio [100%-(5)-(6)]	75.7%
(8) Statewide Projected Average Required Premium [(3)+(4))]/(7)	$ 850.21
(9) Indicated Rate Change (8)/(1)-100%	0.0%

* Countrywide data is used in example.

Exhibit 2-D

Company
State XX Homeowners
Calculation of Proposed Expense Fee
Proposed Method

(1) Total Projected Average Fixed Expense Per Exposure	$	93.28
(2) Variable Expense Provision		19.3%
(3) Profit and Contingencies Provision		5.0%
(4) Proposed Expense Fee [(1)]/[100%-(2)-(3)]	$	123.22

Exhibit 3-A

Comparison of Results

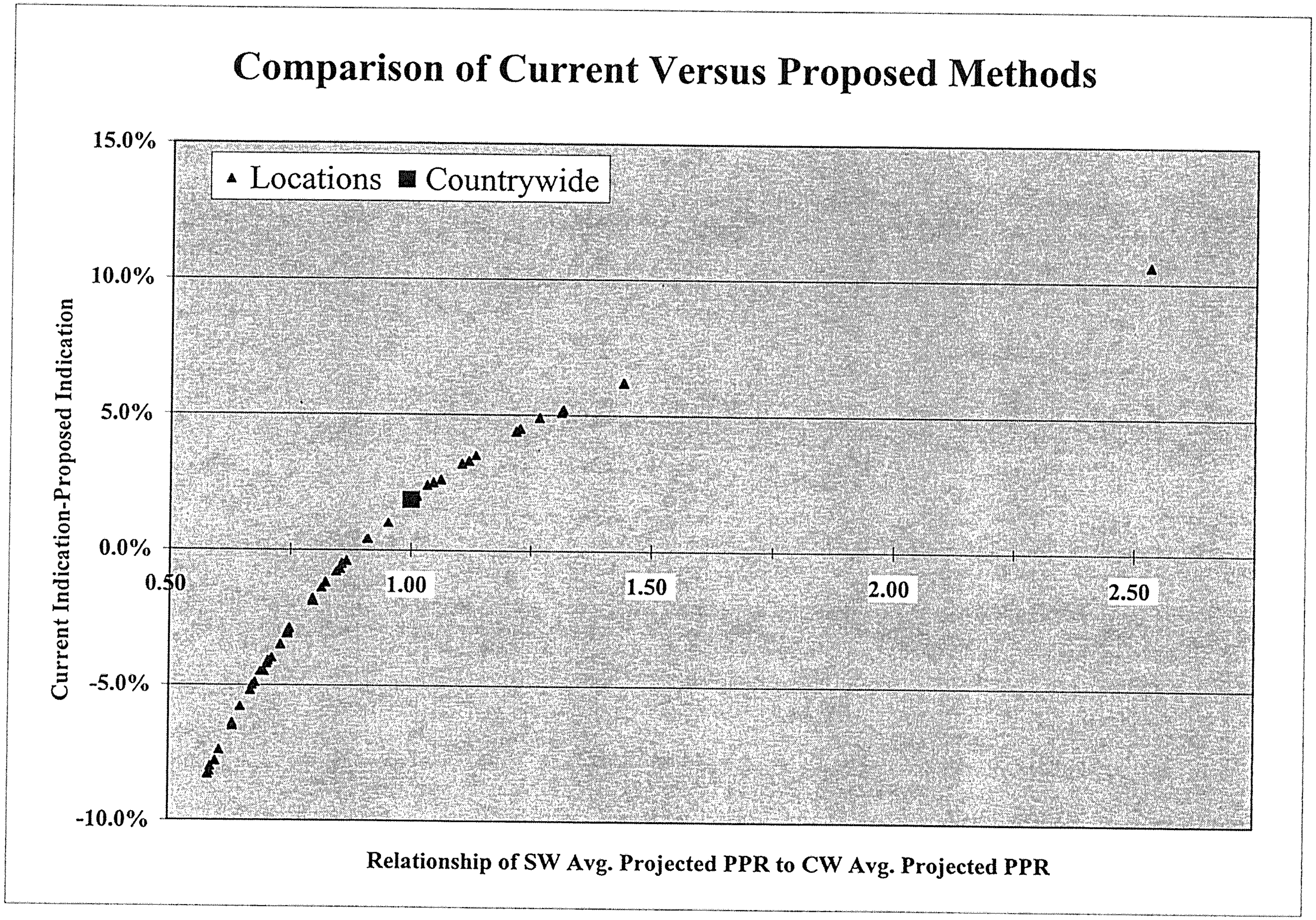
Location	(1) Average Projected Premium at Present Rates Relativitity	(2) Indication* Current Methodology	(3) Indication* Proposed Methodology	(4) Current Indication - Proposed Indication (2)-(3)
1	2.53	1.7%	-8.9%	10.6%
2	1.44	1.3%	-4.9%	6.2%
3	1.44	1.7%	-4.5%	6.2%
4	1.31	1.8%	-3.4%	5.2%
5	1.31	1.3%	-3.8%	5.1%
6	1.27	1.7%	-3.2%	4.9%
7	1.23	1.7%	-2.8%	4.5%
8	1.22	1.2%	-3.2%	4.4%
9	1.13	1.8%	-1.7%	3.5%
10	1.12	1.3%	-2.0%	3.3%
11	1.11	1.5%	-1.7%	3.2%
12	1.06	1.7%	-0.9%	2.6%
13	1.05	1.6%	-0.9%	2.5%
14	1.03	1.7%	-0.7%	2.4%
15	1.01	1.7%	-0.3%	2.0%
16	0.95	0.5%	-0.5%	1.0%
17	0.91	1.2%	0.8%	0.4%
18	0.91	1.2%	0.8%	0.4%
19	0.87	0.7%	1.1%	-0.4%
20	0.86	1.7%	2.2%	-0.5%
21	0.85	1.7%	2.4%	-0.7%
22	0.85	1.7%	2.4%	-0.7%
23	0.85	1.3%	2.1%	-0.8%
24	0.82	2.1%	3.3%	-1.2%
25	0.82	1.7%	2.9%	-1.2%
26	0.81	1.5%	2.9%	-1.4%
27	0.80	0.5%	2.4%	-1.9%
28	0.80	1.1%	2.9%	-1.8%
29	0.80	1.7%	3.6%	-1.9%
30	0.75	1.5%	4.4%	-2.9%
31	0.75	1.7%	4.8%	-3.1%
32	0.75	1.5%	4.5%	-3.0%
33	0.75	1.6%	4.6%	-3.0%
34	0.74	1.7%	4.8%	-3.1%

Exhibit 3-A

Comparison of Results

	(1)	(2)	(3)	(4)
	Average Projected Premium at Present Rates	Indication*		Current Indication - Proposed Indication
Location	Relativitity	Current Methodology	Proposed Methodology	(2)-(3)
35	0.73	1.7%	5.2%	-3.5%
36	0.71	1.3%	5.3%	-4.0%
37	0.71	1.8%	5.9%	-4.1%
38	0.70	1.6%	5.8%	-4.2%
39	0.70	0.8%	5.3%	-4.5%
40	0.69	1.7%	6.2%	-4.5%
41	0.68	1.6%	6.5%	-4.9%
42	0.67	1.2%	6.2%	-5.0%
43	0.67	1.7%	6.9%	-5.2%
44	0.65	1.7%	7.5%	-5.8%
45	0.63	1.7%	8.2%	-6.5%
46	0.63	1.8%	8.2%	-6.4%
47	0.61	1.6%	9.0%	-7.4%
48	0.60	1.2%	9.0%	-7.8%
49	0.59	2.0%	10.0%	-8.0%
50	0.59	1.7%	9.9%	-8.2%
51	0.58	1.6%	9.9%	-8.3%
Countrywide	1.00	1.8%	0.0%	1.8%

* Loss ratio set at 64.7% to make CW indication equal 0% for proposed methodology.


Comparison of Current Versus Proposed Methods
Locations
Countrywide
15.0%
10.0%
5.0%
0.0%
-5.0%
-10.0%
Current Indication-Proposed Indication
0.50
1.00
1.50
2.00
2.50
Relationship of SW Avg. Projected PPR to CW Avg. Projected PPR

A View Inside the "Black Box:" A Review and Analysis of Personal Lines Insurance Credit Scoring Models Filed in the State of Virginia

Cheng-sheng Peter Wu, FCAS, ASA, MAAA, and
John R. Lucker, CISA

A View Inside the "Black Box": A Review and Analysis of Personal Lines Insurance Credit Scoring Models Filed in the State of Virginia

by

Cheng-sheng Peter Wu, FCAS, ASA, MAAA and John R. Lucker, CISA

Abstract

In order to reveal and better understand the inner workings of insurance credit scoring models used by the vast majority of personal lines insurers, the authors obtained nine private passenger automobile and two homeowners' filings from nine insurance groups from the Virginia Bureau of Insurance. Within these filings the authors found three categories of models created by either Fair Isaac & Company, ChoicePoint, or the insurance companies providing the filings. Based on the review and aggregation of these filings, the authors will describe the data sources, scoring functions, scoring algorithms, model variables, and statistical details of these models. In addition to descriptive information, interpretive and explanatory details for the models will be included based on the authors' past experience in conducting predictive modeling projects that included both mainstream and non-traditional predictive variables as well as personal credit information. As a result, the readers will gain a better understanding of how the insurance industry utilizes credit information to formulate insurance credit scores.

About the Authors

Cheng-sheng Peter Wu, F.C.A.S, A.S.A., M.A.A.A., is a director in the Advanced Quantitative Services practice of Deloitte & Touche's Actuarial and Insurance Consulting Group. He is based in the Los Angeles, CA office. Mr. Wu received his Masters degrees in chemical engineering and statistics from the Pennsylvania State University. Mr. Wu has published several papers in automotive engineering, tribology (lubrication engineering), statistics, and actuarial science including two recent CAS and Contingencies articles on insurance credit scoring.

John R. Lucker, CISA, is a senior manager in the Advanced Quantitative Services practice of Deloitte & Touche's Actuarial and Insurance Consulting Group. He is based in the Hartford, CT office. Mr. Lucker received his BA in Biology and his MBA in Marketing and Computer Information Systems from the University of Rochester. Mr. Lucker has spoken on a variety of insurance predictive modeling topics (including insurance credit scoring) and has published several papers on insurance issues. He is also often quoted in insurance trade journals and publications on a variety of insurance industry topics.

Mr. Wu's address is: Deloitte & Touche LLP, 350 South Grand Avenue, Los Angeles, CA 90071 (pwu@deloitte.com) (213-688-5231).

Mr. Lucker's address is: Deloitte & Touche LLP, 185 Asylum Street-33rd Floor, Hartford, CT 06103 (jlucker@deloitte.com) (860-543-7322).

Acknowledgements

The authors would like to thank Hrisanthi "Chris" Adamopolous and Michael Brodnick for their assistance researching topics for this article. We would also like to thank the CAS Ratemaking Committee and our colleagues in the Advanced Quantitative Services (Data Mining and Predictive Modeling) practice at Deloitte & Touche for their review and critique of the form and content of this article.

I. Introductions

I.1 Credit Scoring as a Hot Topic

For nearly 15 years the personal lines P&C insurance industry has been studying and utilizing individual credit history in a variety of ways for ratemaking and underwriting activities. Over this timeframe, the industry's use of credit history as a tool has become extremely widespread to the point that as of 2001, 92% of the respondents to a Conning and Company survey use some form of credit scoring [1]. The respondents to this survey represent approximately 45% of the top 100 insurers by premium volume.

Most of these companies consider the details of the credit scoring models they use, and the methodologies with which they use them, to be proprietary and not something to be publicly disclosed or openly discussed. This "black box" image coupled with a variety of usage and implementation questions and anomalies experienced by consumers and a general concern over personal information privacy, has sparked considerable public debate about the appropriateness of insurance credit scoring [2,3].

Despite this widespread use of credit history in the insurance process, over the past few years the topic of credit scoring as a risk selection and pricing tool for private passenger automobile and homeowner's insurance has become an extremely hot legislative and regulatory topic. The topic has been at the forefront of a variety of newspaper, magazine, television, and radio coverage as well as high profile state legislative and insurance regulatory debates [2,3]. Recently, actuaries have been actively participating in the debate [4,5,6,7] as well.

Much of the concern and doubt may stem from the insurance industry's historical unwillingness to open up the credit scoring "black box" and show consumers, regulators, and industry watchers what is inside. Without a detailed understanding of what's inside the "box" and how it corresponds to insurance risk and policy pricing, allegations have emerged regarding the unfairness of credit scoring and the manner in which it is assumed to discriminate against various population subgroups.

On one side of the debate are the insurance companies and insurance trade organizations that passionately defend the use of credit scores as a valid insurance tool. They argue that the use of credit as a predictor of future insurance loss is statistically valid, objective and impartial, and does not discriminate in favor of, or against, various societal groups.

On the other side are the consumer action groups and other public advocacy organizations. Their numerous concerns include arguments that credit scores are inherently biased (even if that is not the intent of the scores), contain inaccurate information that is difficult to correct, favor one socioeconomic group over another, or do not provide scientifically valid results. They conclude by saying that the use of credit scores as an insurance risk selection and underwriting tool display correlation but not causality.

As a result of the debate and its political sensitivity, state legislators and state insurance departments have stepped up their review and scrutiny of the issue. By Spring 2003, nearly 40 states have introduced new or additional proposals, or have laws pending, for legislative and regulatory control over insurance credit scoring [8]. Furthermore, in 2002 alone, eleven states passed laws that control or limit the use of credit scoring. These laws contain language that ranges from outright prohibitions on credit scoring in specific contexts to controls around the applicability of credit scoring techniques, credit data usage, information disclosure to consumers, risk selection and pricing methodologies, underwriting action protocols, etc.

I.2 The "Black Box" Opens

One of the states focusing considerable attention on the insurance credit scoring issue was Virginia. On June 17, 2002, Alfred W. Gross, Virginia Commissioner of Insurance, issued Administrative Letter 2002-6 mandating that "all insurers licensed to write private passenger automobile insurance and homeowners' insurance in Virginia" must file credit scoring models that are used for risk rating or tiering. Furthermore, Mr. Gross stated, that all such filings would become "part of the [company's] rate filing and will be open to public inspection" according to state law. This action is an example of a recent trend that allows the public more direct scrutiny of the inner workings of insurance credit scores.

I.3 Models Available for Study and How They Are Described in this Paper

Shortly after the filing deadline and using a November 2002 Virginia Bureau of Insurance (DOI) list of forty or so companies that had filed credit scoring models with the Bureau as given in Exhibit 1, we selected and obtained copies of eleven filings, nine for private passenger automobile and two for homeowners', to be used for this study. Our methodology for selecting these models was to obtain representative samples of the different types/categories of models that are used by the population of filed companies for both private passenger automobile and homeowners' coverages.

Upon initial examination, it became readily apparent that the majority of filings utilized "industry" or "customized" models that were developed and provided by either Fair Isaac Company (Fair Isaac) or ChoicePoint. A few large national companies also filed models that were apparently developed by the filing insurers and are "proprietary" to those companies. We therefore obtained filings for companies falling in each of these three model source categories: (1) Fair Isaac; (2) ChoicePoint; and (3) Proprietary.

I.4 What this Study Is and What it is Not

As members of the insurance industry we do not feel it is essential to be too revealing in this paper. We respect the rights of insurance companies to preserve some level of confidentiality and propriety with regards to the credit scoring models that they use and how they use them. As a result we will not be attributing insurance company names to the model details that we will discuss in this paper. We will name Fair Isaac and ChoicePoint as creators of categories of models however because we believe that, as industry standard model providers and vendors of insurance bureau scores, many of the details for their models described in the filings have already been provided in public testimony, public relations descriptions, and are standard and consistent across companies utilizing them. We believe that their models represent the current majority of insurance credit score usage and as such can be openly discussed without targeting proprietary company specific information.

We intend for this paper to review the basic form and structure of insurance credit scoring models. Such information coupled with our analysis should provide additional understanding of how these models combine personal credit history and credit information to formulate a scoring process that is representative of a consumer's general financial stability, credit utilization and behavior, and a consumer's pattern of debt payment. We will not be rendering any judgments on which models are better or worse as business tools for predicting insurance claims experience or insurance policy profitability. Rather we will, as objectively as possible, examine and analyze the filings we have obtained and describe patterns, similarities, differences, strengths, and weaknesses in the structure and logic of the models.

II. Review of the Credit Scoring Models

II.1 General Information for the Reviewed Credit Scoring Models

As stipulated by the Virginia Bureau of Insurance, insurance companies need to file the details of their credit scoring models if they utilize credit scores for rating and tiering in the state. Exhibit 2 lists the eleven filings reviewed in this study, and some of the general information regarding these filings is as follows:

- The eleven filings are from nine insurance groups that were selected from the companies in Exhibit 1. The 2002 homeowners' and private passenger automobile premium written by these 9 insurance groups according to the Best data [9] are also given in Exhibit 2.

- Nine of the eleven filings are private passenger automobile filings, and the other two filings are homeowners' filings. Exhibit 1 shows that there are more credit filings for private passenger automobile than for homeowners' in Virginia, suggesting that credit scores, industry-wide, are used more often in private passenger automobile than in homeowners'. This is consistent with much of the recent industry discussion surrounding the use of credit scores and our experience.

- The models we reviewed fall into three major groups – Industry/Fair Isaac models, Industry/ChoicePoint models, and proprietary models. There are four Fair Isaac models used by three insurance groups, three models for private passenger automobile and one for homeowners'. Another four insurers use the ChoicePoint model for their private passenger automobile books of business. Finally, two insurers developed their own proprietary models. One of them uses one model for both homeowners' and private passenger automobile, while the other uses one model for their private passenger automobile business.

In the following sections, we will discuss these models in detail.

II.2 Industry - Fair Isaac Private Passenger Automobile and Homeowners' Models

Fair Isaac Inc. offers a series of models, called Fair Isaac models, to the marketplace in two primary varieties, private passenger automobile vs. homeowners', and by market segment (e.g. preferred vs. standard vs. non-standard). The models are updated by Fair Isaac on a periodic basis to reflect recent data experience and other data factors. Insurers can pick and choose which model(s) they want to use through discussions with Fair Isaac regarding the predictive power, variable characteristics, etc. for their menu of models. In general, these Fair Isaac models share fairly similar variables and scoring algorithms, and we have seen on occasion that private passenger automobile models are sometimes utilized for homeowners' policy credit scoring, or vice versa.

The following further describes the variables and the scoring algorithm of the models. Additional comments regarding the model comparison and model insight will be given in later sections:

- Data Source – the Fair Isaac models analyzed in this study are based on credit information from TransUnion.

- Variables – the Fair Isaac models use from ten to thirteen credit variables. The variables include the following categories: late payment/past due/delinquent information, derogatory information, bad debt/default/unsatisfactory information, collection information, and other variables. Variables in the last group, the "other" group, contain various pieces of account information, such as the number of accounts, history of accounts, etc., and debt/financial leverage information. Exhibits 2 and 3 compare the variables used in the Fair Isaac models with those of the other models.

- Scoring Algorithm – now that it is available for analysis, the scoring algorithm used in the Fair Isaac models is not overly complex and can be understood in a straightforward fashion. The three steps for the algorithm are:

    1. Assign an individual score for each variable.
    2. Sum the individual score across all variables to derive the total raw score.
    3. Scale the total raw score to the final calibrated score.

- Score Scaling Function - the Fair Isaac models employ a series of linear scaling functions to transform the raw score to the final score by different score ranges. For example, for the Assist 2.0 Fair Isaac model, also labeled as the Preferred Auto Min Limit model, the scaling process is as follows:

- If (raw score+244) < 625 then the final score = raw score + 244
- If (raw score+244) is between 625 and 724 then the final score = 1.2 x (raw score + 244) – 146
- If (raw score+244) >= 725 then the final score = 1.5 x (raw score + 244) – 363.

Please note that the score at the two boundaries within this algorithm (i.e. 390 and 490) is continuous.

- Score Range - the Fair Isaac models assign higher scores to better risks and lower scores to poorer risks. A typical score range is between 200 or 300 for the low end of the score range to 800 for the high end of the score range.

II.3 Industry - ChoicePoint Private Passenger Automobile Model

Unlike the Fair Isaac models, based on our examination of the Virginia filings, ChoicePoint offers the insurance credit scoring marketplace only one model for private passenger automobile and one model for homeowners'. In addition, our historical experience is that ChoicePoint tends to be more willing to provide specific model details for explanatory review and investigation by regulators, insurance companies, and consumers. We reviewed the ChoicePoint model for private passenger automobile in this study.

Compared to other credit models, the ChoicePoint model has the most number of predictive variables and has, perhaps, the most complicated scoring algorithm:

- Data Source – the model in this study uses credit information from Experian.

- Variables – the ChoicePoint model essentially contains two sub-models, one for "thin file" customers with less than four credit accounts on file and the other one for "thick file" customers with four or more credit accounts on file. The model has twenty-nine variables for the "thin file" accounts and thirty-seven variables for the "thick file" accounts. Another unique feature of the ChoicePoint model is that it employs many different types of credit accounts, such as retail accounts, finance accounts, oil and gas card accounts, automotive credit accounts, etc. In addition, ChoicePoint uses many different debt/credit leverage ratio variables, such as the ratio of outstanding balance to available credit limit on open bank revolving accounts. Further study indicates that many variables in this "other" group do not appear to have a significant contribution to the final score.

- Scoring Algorithm – the model's algorithm is perhaps the most complicated algorithm among all of the models reviewed in this study. The algorithm has five steps:

1. Determine whether an account is a "thin file" or a "thick file" account - while the algorithm and structure are the same between the two accounts, the parameters, the variables, and the score scaling functions are different.
2. Calculate ScoreCard 1- in a manner similar to the Fair Isaac algorithm, for ScoreCard 1 the model first assigns a weight to each variable and then sums the individual weight across all variables. However, unlike the Fair Isaac models that use integer scores, ChoicePoint assigns real and fractional numbers to the weights. There are sixteen variables used for the thin file ScoreCard 1 and twenty variables used for the thick file ScoreCard 1.
3. Calculate ScoreCard 2- to calculate the ScoreCard 2 value, the model assigns different weights to each variable, sums the weights across all the variables, and then takes the exponential function of the summed result. It should be noted that exponential functions are widely used for log-linear regression modeling processes, GLM, and neural networks. There are twenty variables used for the "thin file" accounts and fifteen variables used for the "thick file" accounts.
4. Calculate the raw score - the raw score is equal to the ratio of ScoreCard 1 to ScoreCard2. Since ScoreCard1 is a linear function and ScoreCard2 is an exponential function, the ratio of the two will create a unique and complex feature for the raw score. We are not certain as to what the advantages are in employing different functions for the ratio. One thing we notice is that while ScoreCard 2 uses an exponential function, many of the weights are fairly small. When such weights are small in an exponential function, the result can be approximated by a linear function. Another unique feature of the algorithm is that several variables exist in both ScoreCard 1 and ScoreCard 2. Therefore, their overall contribution will be a combination of the contributions in the numerator and in the denominator of the ratio. Due to these complexities, the algorithm is not as easily understood and the effect of each variable to the overall score is not as apparent as other models.
5. The last step is to scale the total raw score to the final calibrated score. While the weights and variables are different between the thin and the thick files, the raw scores are calibrated so that the final scores are brought to the same level between the thin and the thick file policyholders.

- Score Scaling Function – in a manner that is similar to the Fair Isaac models, the ChoicePoint model employs a series of linear scaling functions to transform the raw score to the final score by different score ranges. The model uses different scaling functions between the thin file and the thick file accounts so that the final score gets calibrated to the same level between the two.

- Score Range - as with the Fair Isaac models, the ChoicePoint model also assigns higher scores to better risks. The ChoicePoint score range is, in theory, between 200 and 998.

II.4 Company #1's Proprietary Private Passenger Automobile and Homeowners' Model

The first proprietary model reviewed in this study is applied to both homeowners' and private passenger automobile. The model algorithm and variables appear to be similar to the Fair Isaac models, but the score scaling function and the final score range are different:

- Data Source – the model uses credit information from TransUnion.

- Variables – the model has ten variables. One group of variables used in other models that are not included in this model are the bad debt, account default, and unsatisfactory account information types of variables.

- Scoring Algorithm - the algorithm has three main steps:

    1. Assign an individual score for each variable.
    2. Sum the individual score across all variables to derive the total raw score.
    3. Scale the total raw score to the final calibrated score.

- Score Scaling Function - the transformation process to scale the raw score to the final score is straightforward:

    - Final score = raw score + 100

- Score Range - unlike the industry Fair Isaac and ChoicePoint models, the model assigns higher points to poorer risks and lower points to better risks, and the score range is from 100 to 1000.

II.5 Company #2's Proprietary Private Passenger Automobile Model

The second proprietary model reviewed in this study is for private passenger automobile. The following describes the variables and scoring algorithm used in the model:

- Data Source – the model uses credit information from TransUnion.

- Variables – the model employs many variables, thirty six variables, but several of the variables are transformed in a manner whereby there are multiple timing variables for the same event and occurrence. One example is the number of the accounts open over the past 12, 18, and 24 months. Therefore, there are three variables used for the same information that are different only in the length of the experience period. Another example is the number of accounts 30 or more days past due within the past 3 months, 6 months, 12 months, 18 months, and 24 months. One group of variables not used in this model is data pertaining to collection information.

- Scoring Algorithm – the model employs a linear scoring function with a highly nonlinear score scaling function:

    - Calculate the raw score - the raw score is calculated by a standard linear formula that combines an intercept term and a series of the products of parameters and variables.

    - Transform the total raw score to the final calibrated point.

- Score Scaling Function – the model employs a highly non-linear function that will scale the raw score to a final score range of from 1 to 100:

$$\text{Final Score} = 100 * (1.0056 - \{[8.8533/(\text{Raw Score} + 9.0691)]^{2.3009}\})$$

  Exhibit 4 shows in a graphical format the relationship between the raw score to the final score.

- Score Range: Like the first proprietary model, this model assigns higher scores to poorer quality risk and lower scores to better quality risk. The score range is from 1 to 100.

III. Model Comparison and Additional Comments

Exhibits 2 and 3 summarize some of the model discussion points described above. We will now provide some additional comments regarding the models.

III.1 Model Variables

- In general, the credit variables used in these models can fall into six categories:

    - Late Payment/Past Due/Delinquent information,
    - Derogatory information,
    - Inquiry information,
    - Collection information,
    - Bad Debt/Default/Unsatisfactory information, and
    - Other (such as Debt Leverage information and Account information)

- It appears that all of the models use information for late payment, derogatory, and inquiry information. One model does not use collection variables and one another model does not use variables in the bad debt group. Variables in the last group, other group, vary the most from one model to another.

- Exhibit 5 summarizes how each of the variables impacts the risk quality predictions in these models. For most of the variables, their impacts on the risk quality are consistent among the models. For example, all of the models indicate that higher numbers of late payments are indicative of a poorer quality risk. Few of the variables may exhibit a U-shape or an up and down (or vice versa) shape relationship with regards to the resulting score.

- The variables that differ the most from one model to another are in the "other" group. This is especially true for the ChoicePoint models and Company #2's proprietary model. Each of these models has more than 10 variables in this category. However, we notice that other than a few variables, such as the history of the account, most of the variables in this category do not seem to have a significant contribution to the final score. Perhaps, these variables are used to explain a multivariate effect when the models are applied to different books of business.

- It appears that the variables and the scoring algorithms used in the various Fair Isaac models are quite similar. Between the private passenger automobile and homeowners' models, many of the same variables are utilized. The fact that the Fair Isaac private passenger automobile and homeowners' models are similar suggests that there exist many common underlying credit characteristics between private passenger automobile and homeowners' that are correlated with private passenger automobile and homeowners' losses. Company #1's usage of the same model for both homeowners' and private passenger automobile further supports this indication. However, in order to directly compare different models when they are applied to a book of business, the scores need to be normalized. One approach to achieve such a score normalization is to use score ranking, instead of the final score. This will be explained further in section III.4.

- In order to evaluate the influence of a variable on the model, we can perform a delta method, which is a variable sensitivity test. This method is to evaluate the change in the final score by varying the value of a variable one at a time. The change in the variable values will first impact the raw score, then the final score.

  For illustration purposes, we will show in Exhibit 6 such test for two variables, the average months in the credit file of all accounts and the number of inquiries for the credit file, on three different Fair Isaac models. Exhibit 6 shows that for the Assist 2.0 – Preferred Auto Min Limits model, for example, when average months in file changes from 0-20 months to 21-23 months, the final credit score will increase by "1" for total score <625, by "1" for score between 625 and 724, and by "2" for score >=724. It is interesting to point out that for the two variables tested in Exhibit 6, how the variables impact the final credit score is the same among the three models - the higher the average months in file, the higher the score and the better the risk; the higher the number of inquriries, the lower the score and the worse the risk. Next, we perform a normalization test to compare the degree of the impacts. Due to different score ranges and scaling functions used from one model to another, we need a normalization process to have a meaningful comparison, and the result is given in Exhibit 7. In Exhibit 7, for the impacts in each column from Exhibit 6, we divide them by the first number in the column. Exhibit 7 clearly indicates that for the two variables tested here, their strengths are very similar among the three models.

Two more things need to be noted for the variable sensitivity test:

1. For a non-linear scoring formula such as the ChoicePoint approach, the result is not a constant and is dependent on the values of other variables.

2. Sometimes multiple variables are correlated and they are not completely independent. For example, for the Company #2's proprietary model, there are three late payment variables, late payments over the past 12, 18, and 24 months. Therefore, if we increase the late payment variable from 0 to 1 for the test, then we need to increase the other 2 late payment variables to 1 as well.

III.2 Scoring Algorithm

- In general, the scoring approaches fall into two main categories – the rule-based approach and the formula approach.

- Rule-based scoring approach – this approach refers to the algorithms that assign points or scores directly to each variable. Such algorithms in fact create a series of "if-then rules" to determine the final credit score. The advantages and disadvantages of this approach are:

    - It is an approach that is relatively simple, easy to understand, and easy to communicate and explain.
    - This approach produces a rating mechanism that can fit in well with the insurance rating and class plan structure. Therefore, it can be easily incorporated into the rating and class plan reviews. For example, actuaries can use minimum bias [10,11,12] or GLM technique [13,14] to determine the factors for each "rule" the underlies the credit variables along with the rating variables in a comprehensive class plan factor analysis.
    - One disadvantage of this approach includes the significant effort that is required to pre-determine the groupings for the rules in each variable.
    - Another disadvantage is the potential for low credibility if the number of groupings and variables increases, leading to high volatility in the results. Therefore, there are less variables used in the approach than the formula approach, which will be described next.

- Formula approach – this approach determines the score through a mathematical formula. Therefore, the key information for the approach is the determination of the weights that apply to each of the variables. The advantages and disadvantages of the approach are:

- Most of the modeling techniques, including regression, GLM [13,14], neural networks [15,16,17,18], and MARS [19,20], create formulas directly. It is therefore easier for modelers to apply the modeling results through the formula approach.
- It is easier for this approach to utilize more variables than the rule-based approach.
- One major disadvantage of this approach is that the resulting formulas are more complex and more difficult to understand by reviewers who may not have backgrounds in actuarial science and statistics. Even linear-type formulas can be complex and lengthy. Therefore, this approach often creates a "black-box" mentality around the models, especially for more complicated model generation techniques that may be utilized such as neural networks.
- Another disadvantage of the formula approach is that when more and more variables are included in the models, on occasion, the weights/coefficients to some variables become difficult to interpret in relationship to their business context and meaning. For example, formula coefficients can become counter-intuitive when a variable is expected to indicate a poorer quality risk but the model's coefficient seems to indicate the opposite. The model's complicated mathematical interactions between the variables and their related coefficients often cause such conditions. Such formula characteristics can create a challenge for companies using such models in today's regulatory environment. It is difficult to explain to regulators and consumers the mathematical basis for such model structures.

- One way to connect the two approaches (rule-based and formula) is through the delta method that we have previously described. Since the rule-based approach will assign points for each value of a credit variable, then the difference in points between two adjacent values of a variable is the delta change. Such delta changes can be derived from the formula approach as well, which will be the same as the weights for a linear formula, or the first derivative for a non-linear formula. By comparing the delta between these two approaches, the results of the two approaches can be connected or compared.

- Among the models reviewed in this study, the Fair Isaac and Company #1's proprietary models utilize the rule-based approach, while the Company #2's proprietary model employs the formula approach. Interestingly, ChoicePoint's model is a mixed approach of both profile and formulas. The ChoicePoint model presents the ScoreCard1 and ScoreCard2 as a rule-based point-assignment format, but the process appears to be more similar to the formula approach.

III.3 Score Scaling Functions and Final Score Range

- The purpose of the score scaling functions is to transform a raw score, which is typically a continuous and fractional number that predicts loss ratio, to a final score that can be used and understood more easily by end users, including underwriters, agents, publics, and regulators.

- The scaling functions have to be monotonic – that is, strictly increasing or decreasing so that one unique value for the raw score will be transformed into a unique final score.

- The selection of a scaling function is determined by the final range and the distribution of the score in the range. The scaling function can also be influenced by other criteria. An example is to benchmark the score to loss ratio relativity such as score 200 for –30% relativity, 300 for –20%, 400 for –10%, 500 for 0%,...etc.

- The scaling functions used by Fair Isaac, ChoicePoint, and Company #1's proprietary model are simple linear scaling functions, while the scaling functions used by Company #2's proprietary model are highly nonlinear.

- It is interesting that the two industry models, ChoicePoint and Fair Isaac assign higher scores to better risks, while the two proprietary models assign higher scores to worse risks. We assume that the difference has more to do with how the scores are utilized within the company's information systems than with any other factor.

III.4 Normalization Score Ranking Testing with Real Data

- In previous sections, we commented on the details of each model with regard to their variables, scoring algorithm, score scaling functions, and final calibrated score range. We also discussed a normalization process with the variable sensitivity test to compare variables' strength between models. The ultimate comparison of the models is on the final score. While there exists many similar characteristics among these models, such as the variables used in the models, they differ in areas such as scoring algorithms and score ranges. The best way to compare and evaluate various models is to test them with real data. Our experience indicates that the model results are highly influenced by the market segmentation (preferred vs. standard vs. nonstandard), demographic distribution (i.e. age, gender, marital status, etc), and geographic distribution (such as urban vs. suburban vs. rural).

- Since the model variables, scoring functions, and score ranges vary from one model to another for a book of business, a normalization process of using score ranking can be used when comparing different models. This normalization process includes the following steps: (1) score the data set; (2) sort the data points from best to worst based upon their predicted outcomes; and (3) assign the score to each data point based on the relative ranking of the predictive result. The resulting predictive model scores can then be sorted into a fixed percentile range or a range of buckets, for example, 10 equal size buckets or deciles. This process is called a decile analysis (or 5 equal size bukets for quintles, 4 equal size bukets for quartiles, etc).

- To illustrate the score ranking test, we have scored a real dataset of 10,000 data points using two Fair Isaac models – the Fair Isaac - Assist 2.0 Preferred Auto model and the Fair Isaac - Assist 2.1 HO3 model. First, we score the data by the individual model and Exhibit 8 shows the score distribution by the two models. Exhibit 9 then compares the difference in the scores between the two models. At first glance, Exhibits 8 and 9 may appear to show that the difference in the score between the two models is quite significant. However, if we rescale the final score into a decile score ranking for comparison, the results indicate that the difference between these two models is not significant. This is illustrated in Exhibit 10 where, for example, it is apparent that almost 90% of the data points have a differential within +-3 deciles between the two models.

  In addition to comparing the score disruption from one model to another, we can also use this score ranking process to test how each underlying variable affects the model's final outcomes through the delta method described previously. Instead of testing the impact on the score change as given in Exhibit 6 and Exhibit 7 we can test the impact of the variable on the score ranking change.

IV. Consideration of Selecting a Credit Score Model

When a company evaluates whether to build a proprietary credit model or select an off-the-shelf industry model, we believe that the following issues should be carefully evaluated:

- Predictive Power/Lift: The name of the game is to build a model that has strong predictive power and the ability to maximize the segmentation of better risks from poorer risks. A standard measurement that can be used to represent such predictive power and segmentation capability is the concept of a "lift" curve [5]. A lift curve is generated by sorting the score of a test set, breaking the dataset into equal-sized pieces (for example, 10 pieces/deciles), and then plotting the loss ratio/frequency/severity for each piece. We have portrayed a lift curve in Exhibit 11 using deciles as the unit of division. If a model can successfully segment better risks from poorer risks, then the curve should exhibit an increasing slope from better deciles to poorer deciles. The higher the slope, the more predictive power the model possesses. A series of benchmarks can be derived from the lift curves, which includes the loss ratio relativity for the best and the worse 5%, 10%, and 25%. Another commonly used benchmark is the ratio of the loss ratio for the worst 10% to the loss ratio of the best 10%, or the worst 25% to the best 25%. For example, in Company #1's filing, it indicates a surcharge of 45% for the worst 10% and a discount of 25% for the best 10%. This suggests a ratio of 1.45/0.75 = 1.8 for the worst 10% risks to the best 10% risk. Our experience indicates that in general the ratio of worst to best 10% for a typical credit score model is between 1.5 to 2.0

- Stability of a credit model: Another consideration when evaluating the quality of a credit scoring predictive model is the model stability, or how frequently and by how much a credit score will change from one period to another. This consideration is particularly important if the credit score is used for renewal pricing. To test the stability of a credit score model one must perform a multiyear analysis. Again, we recommend that such analyses be done using the score ranking test described previously.

- Company expertise and regulatory defensibility: Regardless of the type of model used, an insurer must be able to authoritatively speak about the model and be prepared to defend the model to customers and regulators. The degree to which a company can develop expertise about the inner workings of a model depends somewhat on whether the model is a vended model or a custom proprietary model. A custom proprietary model allows a company to better control the way in which the model is designed, developed, and implemented. And the mere fact that the model is the property of the company tends to make it more likely that the company will have greater expertise and insight into the model and its subtleties.

V. Conclusion: Is It Still a "Black Box"?

In recent years there has been a significant "tug of war" between the insurance industry and insurance regulators/consumers. The "black box" concept fueled the flames. By taking the time to examine the contents of these recent insurance filings, we have been able to gain a better understanding of the techniques and methods of the insurance industry for credit scoring. We have seen that there are more similarities between models than there are differences. Many key variables used in these models are the same. Some scores are manifested in rule-based methods while others are the result of a multivariate formula. While algorithms, scaling functions, and score ranges may vary between the models, in the end, the score represents an assessment of whether a risk is expected to be a better risk or a poorer risk with personal credit data as the primary driver for this indication.

As we have seen, most private passenger automobile and homeowner's insurers use credit scoring as a risk selection and pricing tool. Virtually no insurers use credit scoring in a vacuum; that is, credit scoring is only part of the entire process and not the sole determining factor. Given this supporting role in the insurance process and the significant evidence that insurance credit scoring works for predicting the propensity for insurance loss, it seems unlikely that insurance credit scoring will be banned throughout the country. Instead model legislation like that proposed by the National Conference of Insurance Legislators (NCOIL) seems likely to continue to serve as the basis which will provide fairness to consumers while allowing the insurance industry to use a strong predictive tool. By opening the "black box" through public discussion and analysis similar to that which is presented in this article, the public can become more comfortable with the inner workings of these models and the techniques that are utilized within them. We believe that such understanding will improve the comfort level for the use of insurance credit scores and in the end, openness and dialogue should help resolve some of the differences surrounding the issue.

References

1. "Insurance Scoring in Private Passenger Automobile Insurance – Breaking the Silence", *Conning Report,* Conning, (2001).

2. "Insurers Battling Credit-Scoring", *National Underwriter*, March 5th Issue, (2002).

3. "Insurers Lose a Credit Scoring Battle", *National Underwriter*, February 21st Issue, (2002).

4. Monaghan, J. E., "The Impact of Personal Credit History on Loss Performance in Personal Lines", *CAS Forum*, Casualty Actuarial Society, (2000).

5. Wu, C. P., Guszcza, J., "Does Credit Score Really Explain Insurance Losses? - Multivariate Analysis from a Data Mining Point of View," *2003 CAS WinterForum*, Casualty Actuarial Society (2003).

6. Guszcza, J., Wu, C. P., "Mining the Most from Credit and Non-Credit Data," *Contingencies*, American Academy of Actuaries, March/April (2003)

7. "Report on the Use of Credit History for Personal Line of Insurance," American Academy of Actuaries, (2002)

8. National Association of Mutual Insurance Companies, "State Laws Governing the Use of Credit-Based Insurance Scoring", April 9, 2003.

9. *Best's Aggregates & Averages*, AM Best Company, (2002)

10. Bailey, R. A., "Insurance Rates with Minimum Bias", *Proceedings of Casualty Actuarial Society,* Vol. L, Casualty Actuarial Society, (1960).

11. Bailey, R. A. and Simon, L. J., "Two Studies in Auto Insurance Ratemaking", *Proceedings of Casualty Actuarial Society,* Vol. XLVII, Casualty Actuarial Society, (1963).

12. Brown, R. L., "Minimum Bias with Generalized Linear Models", *Proceedings of Casualty Actuarial Society,* Vol. LXXV, Casualty Actuarial Society, (1988).

13. Mildenhall, S.J., "A Systematic Relationship Between Minimum Bias and Generalized Linear Models," *Proceedings of Casualty Actuarial Society,* Vol. LXXXVI, Casualty Actuarial Society, (1999).

14. Holler, K.D., Sommer, D.B.; and Trahair, G., "Something Old, Something New in Classification Ratemaking with a Novel Use of GLMS for Credit Insurance," *CAS Forum*, Casualty Actuarial Society, (1999).

15. Francis, L. A., "Neural Networks Demystified," *CAS Forum*, Casualty Actuarial Society, (2001)

16. Wu, C. P., "Artificial Neural Networks – the Next Generation of Regression" *Actuarial Review,* Casualty Actuarial Society, Vol. 22, No. 4, (1995).

17. Zizzamia, F., Wu, C. P., "Driven by Data: Making Sense Out of Neural Networks", *Contingencies*, American Academy of Actuaries, May/June, (1998).

18. Dugas, C., Bengio, D., Chapados, N., Vicent, P., Denoncourt, G., "Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking," *2003 CAS Winter Forum*, Casuatly Actuarial Society, (2003)

19. Friedman, J. H., "Multivariate Adaptive Regression Splines," *Annals of Statistics*, Vol. 19, (1991).

20. Francis, L. A., "Martian Chronicles: Is MARS Better than Neural Networks?," *2003 CAS Winter Forum*, Casualty Actuarial Society, (2003)

EXHIBITS

Exhibit 1
List of Insurers Using Credit Scores in Rating - Virginia Bureau of Insurance

This list was delivered to the authors as a draft document as of November 2002. It was obtained from the Virginia Bureau of Insurance and has not been formally verified.
The list may include the names of insurers who no longer use credit scoring in rating or may not include the names of insurers that have recently begun to use credit scoring.

Company Name	Line(s) of Insurance
AIG Group (AIG National Insurance Company; AIU Insurance Company; American Home Assurance Company; American International South Insurance Company)	Personal Auto
Agency Insurance Company of Maryland	Personal Auto
Allstate Insurance Company	Personal Auto and Homeowners
Allstate Indemnity Company	Personal Auto and Homeowners
American & Foreign Insurance Company	Homeowners
American Motorists Insurance Company	Personal Auto and Homeowners
Auto-Owners Insurance Company	Personal Auto - Pending
Deerbrook Insurance Company	Personal Auto
Farmers Insurance Exchange	Personal Auto and Homeowners
First Liberty Insurance Corporation	Personal Auto
Globe Indemnity Company	Personal Auto
Harleysville Mutual Insurance Company	Homeowners
Harleysville Preferred Insurance Company	Homeowners
Hartford Accident and Indemnity Company	Personal Auto
Homesite Insurance Company	Homeowners
Horace Mann Insurance Company	Personal Auto and Homeowners
Horace Mann Property and Casualty Ins Company	Personal Auto and Homeowners
Integon National Insurance Company	Personal Auto
Kansas City Fire and Marine Insurance Company	Personal Auto and Homeowners (USP)
Kemper Auto and Home Insurance Company	Personal Auto
Kemper Independence Insurance Company	Personal Auto
Liberty Mutual Fire Insurance Company	Personal Auto
Main Street America Insurance Company	Personal Auto
Mercury Casualty Company	Personal Auto
Metropolitan General Insurance Company	Personal Auto
Mid-Century Insurance Company	Personal Auto
Montgomery Ward Insurance Company	Personal Auto
National Grange Mutual	Personal Auto
Nationwide Mutual Fire Insurance Company	Personal Auto
Nationwide Mutual Insurance Company	Personal Auto
Owners Insurance Company	Personal Auto
Prudential General Insurance Company	Personal Auto
Royal Indemnity Company	Homeowners
Royal Insurance Company of America	Personal Auto
Safeguard Insurance Company	Personal Auto
State Auto Property and Casualty Ins. Company	Personal Auto
Teachers Insurance Company	Personal Auto
Tri-State Insurance Company	Personal Auto
United Services Automobile Association	Personal Auto and Homeowners
USAA Casualty Insurance Company	Personal Auto and Homeowners

Exhibit 2
Summary of Credit Score Models Reviewed

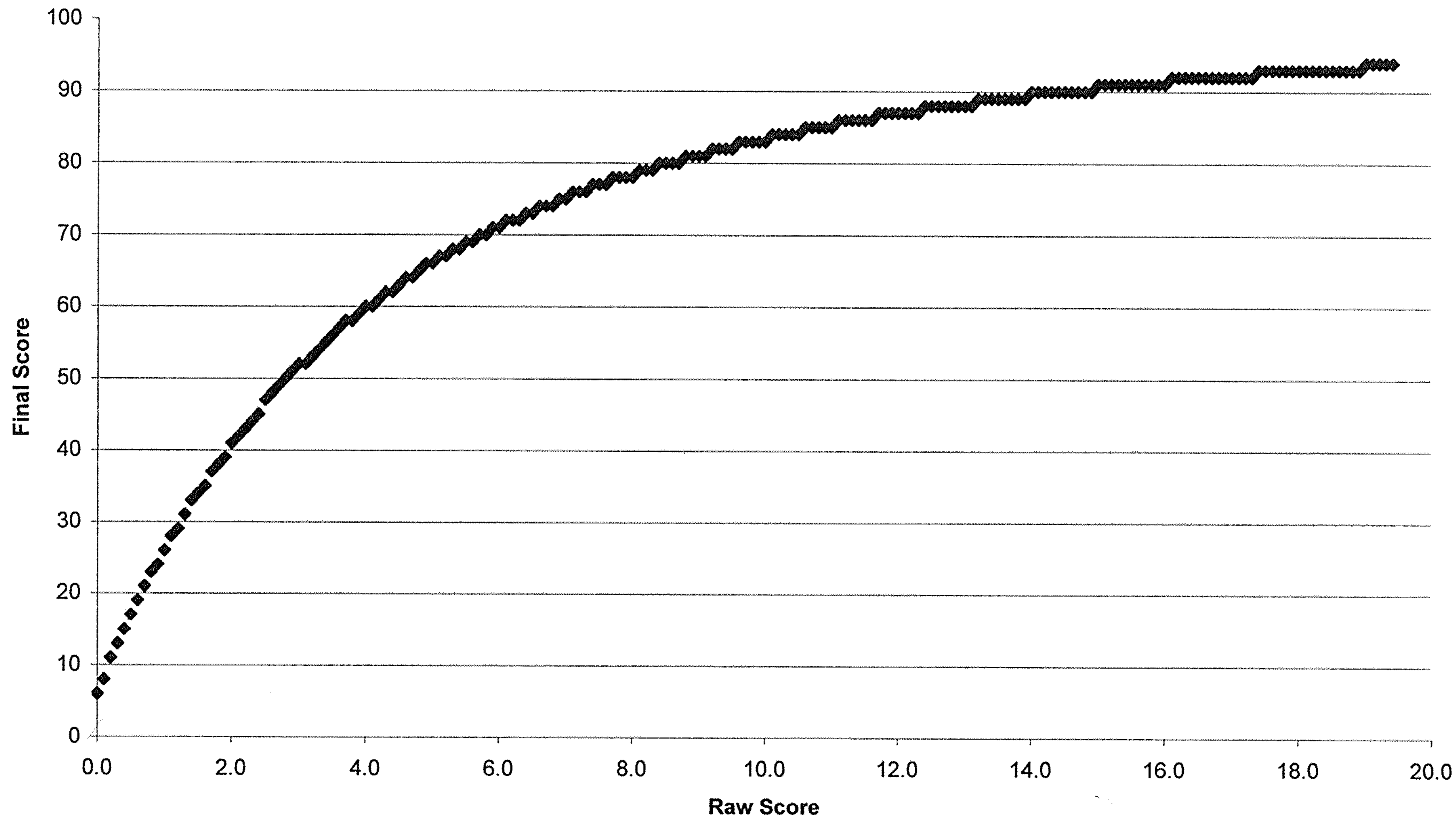
Company	LOB	2002 Net Written Premium Derived from AM Best [8]	Industry vs Proprietary Models	# of Variables	Scoring Functions	Score Range*	Higher Score
Company #1	Auto	>=$5 Billions	Proprietary	10	Profiling	84-1078	Worst Risks
Company #1	Home	>=$5 Billions	Proprietary - Same as Auto	10	Profiling	84-1078	Worst Risks
Company #2	Auto	>=$5 Billions	Proprietary	36	Scoring	1-100	Worst Risks
Company #3	Auto	<=$1 Billion	Fair Isaac - Assist 2.0, Preferred Greater than Min Limits	12	Profiling	357-818	Better Risks
Company #4	Auto	>$1 and <$5 Billions	Fair Isaac - InScore 2.0, Standard Greater than Min Limits	11	Profiling	389-806	Better Risks
Company #5	Auto	<=$1 Billion	ChoicePoint	29 for Thin and 37 for Thick	Scoring	220-997	Better Risks
Company #6	Auto	>$1 and <$5 Billions	ChoicePoint	29 for Thin and 37 for Thick	Scoring	220-997	Better Risks
Company #7	Auto	<=$1 Billion	ChoicePoint	29 for Thin and 37 for Thick	Scoring	220-997	Better Risks
Company #8	Auto	>$1 and <$5 Billions	ChoicePoint	29 for Thin and 37 for Thick	Scoring	220-997	Better Risks
Company #9	Auto	>=$5 Billions	Fair Isaac - Assist 2.0, Preferred Auto Min Limits	11	Profiling	326-845	Better Risks
Company #9	Home	>=$5 Billions	Fair Isaac - Assist 2.1, HO3	13	Profiling	200-884	Better Risks

* The score ranges given here are the theorical score ranges, not the likely score ranges for a typical book of business.

Exhibit 3
Number of Credit Model Variables in Each Group

Company	LOB	Industry vs Proprietary Models	(1) Late Payment/Past Due/Deliquent Information	(2) Unsatifactory, Default, Bad Debt Info	(3) Public Derogatory Information	(4) Collection Information	(5) Inquiry Information	(6) Other - Account Informatio, Levarage Ratio, and Others
Company #1	Auto	Proprietary	3	0	1	1	1	4
Company #1	Home	Proprietary - Same as Auto	3	0	1	1	1	4
Company #2	Auto	Proprietary	17	1	1	0	3	14
Company #3	Auto	Fair Isaac - Assist 2.0, Preferred Greater than Min Limits	5	2	2	1	1	3
Company #4	Auto	Fair Isaac - InScore 2.0, Standard	3	4	1	1	0	2
Company #5	Auto	ChoicePoint	2 for Thin and 4 for Thick	7 for Thin and 6 for Thick	2 For Thin	1 for Thick	2 for Thin and 2 for Thick	10+ for both Thin and Thick
Company #6	Auto	ChoicePoint	2 for Thin and 4 for Thick	7 for Thin and 6 for Thick	2 For Thin	1 for Thick	2 for Thin and 2 for Thick	10+ for both Thin and Thick
Company #7	Auto	ChoicePoint	2 for Thin and 4 for Thick	7 for Thin and 6 for Thick	2 For Thin	1 for Thick	2 for Thin and 2 for Thick	10+ for both Thin and Thick
Company #8	Auto	ChoicePoint	2 for Thin and 4 for Thick	7 for Thin and 6 for Thick	2 For Thin	1 for Thick	2 for Thin and 2 for Thick	10+ for both Thin and Thick
Company #9	Auto	Fair Isaac - Assist 2.0, Preferred Auto Min Limits	4	2	1	1	1	2
Company #9	Home	Fair Isaac - Assist 2.1, HO3	3	2	2	2	1	3

Exhibit 4
Company #2 Proprietary Model's
Score Scaling Function



Pages 275-282 have been removed.
For Exhibit 4 raw data, see the printed publication.

Exhibit 5
Impact of Variables on Credit Scores Suggested by the Models

Variables	More	Recent
Late Payment/Past Due/Deliquent Information	Worse	Worse
Unsatifactory, Default, Bad Debt Information	Worse	N/A
Public Derogatory Information	Worse	Worse
Collection Information	Worse	Worse
Inquiry Information	Worse	N/A
Other - Account Informatio, Levarage Ratio, and Others	Varies	

Exhibit 6
An Example of Testing Variable Strength with the Delta Method

I. Change in Average Months in File of All Accounts

Change in the Final Score

Average Months Change		Fair Isaac - Assist 2.0, Preferred Auto Min Limits			Fair Isaac - Assist 2.0, Preferred Greater than Min Limits		Fair Isaac - Assist 2.1, HO3	
From	To	Score < 625	Score <725 and >=625	Score >=725	Score < 730	Score >= 730	Score <775	Score >= 775
0-20	21-23	1	1	2	2	5	2	3
21-23	24-29	4	5	6	6	15	10	14
24-29	30-32	6	7	9	9	20	12	17
30-32	33-39	1	1	2	2	5	2	3
33-39	40-41	1	1	2	2	5	3	4
40-41	42-47	1	1	2	1	2	2	3
42-47	48-53	8	10	12	14	32	19	26
48-53	54-59	4	5	6	8	17	10	14
54-59	60-65	2	2	3	3	7	5	7
60-65	66-71	1	1	2	2	5	2	3
66-71	72-83	4	5	6	5	12	7	10
72-83	84-89	1	1	2	1	2	3	4
84-89	90-95	1	1	2	2	5	2	3
90-95	96-105	1	1	2	2	5	3	4
96-105	106-115	1	1	2	1	2	2	3
106-115	116-119	1	1	2	2	5	2	3
116-119	120-139	1	1	2	1	2	3	4
120-139	140-159	1	1	2	2	5	2	3
140-159	160-179	1	1	2	2	5	3	4
160-179	180-199	1	1	2	1	2	2	3
180-199	200-219	1	1	2	2	5	2	3
200-219	220-239	1	1	2	2	5	3	4
220-239	240-359	1	1	2	1	2	2	3
240-359	360-479	1	1	2	2	5	3	4
360-479	480-599	1	1	2	1	2	2	3
480-599	600+	1	1	2	2	5	2	3

II. Change in Number of Inquiries

Change in the Final Score

Number of Inquiries		Fair Isaac - Assist 2.0, Preferred Auto Min Limits			Fair Isaac - Assist 2.0, Preferred Greater than Min Limits		Fair Isaac - Assist 2.1, HO3	
From	To	Score < 625	Score <725 and >=625	Score >=725	Score < 730	Score >= 730	Score <775	Score >= 775
0	1	-7	-8	-11	-5	-7	-8	-17
1	2	-12	-14	-18	-11	-15	-13	-29
2	3	-15	-18	-23	-13	-18	-17	-39
3	4	-16	-19	-24	-14	-19	-18	-41
4	5+	-6	-7	-9	-5	-7	-6	-15

Exhibit 7
Normalization of the Testing Variable Strength Results in Exhibit 6 *

I. Change in Average Months in File of All Accounts

Ratio of the Changes in the Final Score Using the First Change as the Base *

Average Months Change			Fair Isaac - Assist 2.0, Preferred Auto Min Limits			Fair Isaac - Assist 2.0, Preferred Greater than Min Limits		Fair Isaac - Assist 2.1, HO3	
From	To		Score < 625	Score <725 and >=625	Score >=725	Score < 730	Score >= 730	Score <775	Score >= 775
0-20	21-23	**	***1.0***	***1.0***	***1.0***	***1.0***	***1.0***	***1.0***	***1.0***
21-23	24-29		4.0	5.0	3.0	3.0	3.0	5.0	4.7
24-29	30-32		6.0	7.0	4.5	4.5	4.0	6.0	5.7
30-32	33-39		1.0	1.0	1.0	1.0	1.0	1.0	1.0
33-39	40-41		1.0	1.0	1.0	1.0	1.0	1.5	1.3
40-41	42-47		1.0	1.0	1.0	0.5	0.4	1.0	1.0
42-47	48-53		8.0	10.0	6.0	7.0	6.4	9.5	8.7
48-53	54-59		4.0	5.0	3.0	4.0	3.4	5.0	4.7
54-59	60-65		2.0	2.0	1.5	1.5	1.4	2.5	2.3
60-65	66-71		1.0	1.0	1.0	1.0	1.0	1.0	1.0
66-71	72-83		4.0	5.0	3.0	2.5	2.4	3.5	3.3
72-83	84-89		1.0	1.0	1.0	0.5	0.4	1.5	1.3
84-89	90-95		1.0	1.0	1.0	1.0	1.0	1.0	1.0
90-95	96-105		1.0	1.0	1.0	1.0	1.0	1.5	1.3
96-105	106-115		1.0	1.0	1.0	0.5	0.4	1.0	1.0
106-115	116-119		1.0	1.0	1.0	1.0	1.0	1.0	1.0
116-119	120-139		1.0	1.0	1.0	0.5	0.4	1.5	1.3
120-139	140-159		1.0	1.0	1.0	1.0	1.0	1.0	1.0
140-159	160-179		1.0	1.0	1.0	1.0	1.0	1.5	1.3
160-179	180-199		1.0	1.0	1.0	0.5	0.4	1.0	1.0
180-199	200-219		1.0	1.0	1.0	1.0	1.0	1.0	1.0
200-219	220-239		1.0	1.0	1.0	1.0	1.0	1.5	1.3
220-239	240-359		1.0	1.0	1.0	0.5	0.4	1.0	1.0
240-359	360-479		1.0	1.0	1.0	1.0	1.0	1.5	1.3
360-479	480-599		1.0	1.0	1.0	0.5	0.4	1.0	1.0
480-599	600+		1.0	1.0	1.0	1.0	1.0	1.0	1.0

II. Change in Number of Inquiries

Ratio of the Changes in the Final Score Using the First Change as the Base *

Number of Inquiries			Fair Isaac - Assist 2.0, Preferred Auto Min Limits			Fair Isaac - Assist 2.0, Preferred Greater than Min Limits		Fair Isaac - Assist 2.1, HO3	
From	To		Score < 625	Score <725 and >=625	Score >=725	Score < 730	Score >= 730	Score <775	Score >= 775
0	1	**	***-1.0***	***-1.0***	***-1.0***	***-1.0***	***-1.0***	***-1.0***	***-1.0***
1	2		-1.7	-1.8	-1.6	-2.2	-2.1	-1.6	-1.7
2	3		-2.1	-2.3	-2.1	-2.6	-2.6	-2.1	-2.3
3	4		-2.3	-2.4	-2.2	-2.8	-2.7	-2.3	-2.4
4	5+		-0.9	-0.9	-0.8	-1.0	-1.0	-0.8	-0.9

* Results given in this exhibit is the ratio of each change to the first change in the same column

** This row is used as the base to normalize the changes.

Exhibit 8
Score Distribution Comparison between a Auto Model and a Home Model by Fair Isaac

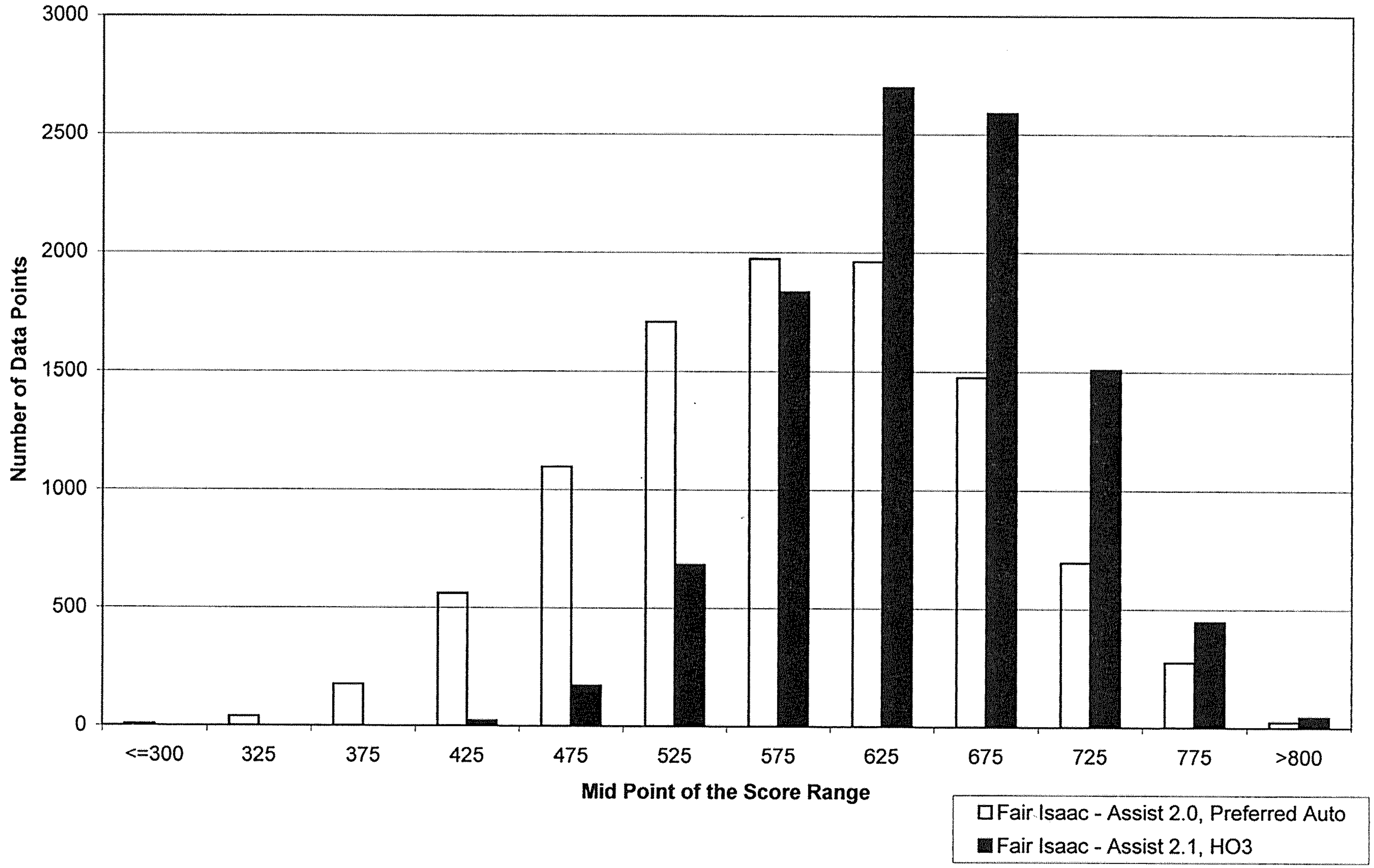


Exhibit 9
Distribution of the Difference Between a Fair Isaac Auto Score and a Fair Isaac Home Score

	Mid Point of the Difference*	Number of Data Points	% of Data Points
	-175	2	0.02%
	-125	185	1.85%
	-75	1452	14.52%
	-25	2979	29.79%
	25	2906	29.06%
	75	1861	18.61%
	125	600	6.00%
	175	15	0.15%
Total		10000	100.00%

* Difference = (Fair Isaac Assist 2.1, HO3 Score) - (Fair Isaac Assist 2.0, Preferred Auto Score)

Exhibit 10

Distribution of the Difference Between a Fair Isaac Auto Score and a Fair Isaac Home Score in Decile Ranking

	Difference in Decile Score Ranking	Number of Data Points	% of Data Points
	-9	0	0.0%
	-8	0	0.0%
	-7	1	0.0%
	-6	23	0.2%
	-5	99	1.0%
	-4	259	2.6%
	-3	599	6.0%
	-2	1068	10.7%
	-1	1646	16.5%
	0	2702	27.0%
	1	1569	15.7%
	2	1024	10.2%
	3	557	5.6%
	4	318	3.2%
	5	115	1.2%
	6	20	0.2%
	7	0	0.0%
	8	0	0.0%
	9	0	0.0%
Total		10000	100.0%

* Difference = (Fair Isaac Assist 2.1, HO3 Score Ranking) - (Fair Isaac Assist 2.0, Preferred Auto Score Ranking)

Exhibit 11

A Loss Ratio Lift Curve

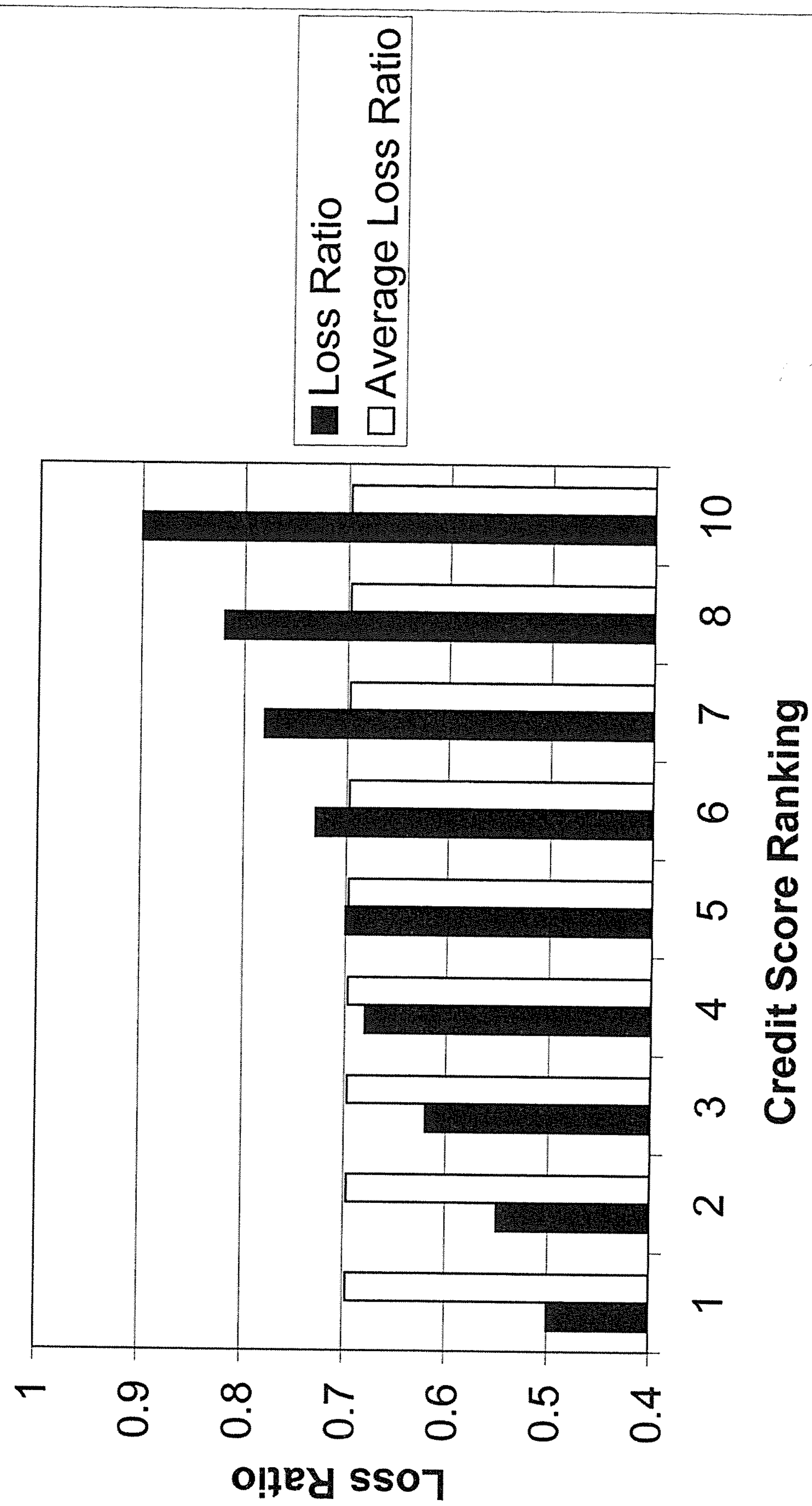


Exhibit 11
A Loss Ratio Lift Curve
Data

Score Ranking	Loss Ratio	Average Loss Ratio
1	0.5	0.697
2	0.55	0.697
3	0.62	0.697
4	0.68	0.697
5	0.7	0.697
6	0.73	0.697
7	0.78	0.697
8	0.82	0.697
10	0.9	0.697

Valuing Stochastic Cash Flows: A Thought Experiment

Leigh J. Halliwell, FCAS, MAAA

Valuing Stochastic Cash Flows: A Thought Experiment

Leigh J. Halliwell, FCAS, MAAA

How should we value stochastic cash flows? Without a common framework our discussions generate more heat than light. The following thought experiment seeks to filter out the extraneous and to focus on the essential.

Imagine that you wish to value stochastic cash flow C that has n possible outcomes. The i^{th} outcome is the receipt of x_i dollars t_i years from now, and the probability of the outcome is p_i. Hence, $\sum_{i=1}^{n} p_i = 1$. The value of C, $V[C]$, is a dollar value, and you would be willing to pay for C as much as, but not more than, $V[C]$. Really, it matters not whether your beliefs about n, x, t, and p are the "right" ones; indeed, others might disagree with you. What matters is that you are willing to act on your beliefs.

Take for granted that you can value deterministic cash flows, i.e., that you can value the guaranteed receipt of x dollars t years from now. Let the function $v(x, t)$ denote that value. For stochastic cash flow C, with $n > 1$ and all $p_i > 0$,[1] your value should lie within the range of the deterministic values; in symbols:

$$\min_i \{v(x_i, t_i)\} \le V[C] \le \max_i \{v(x_i, t_i)\}$$

How could anyone dispute this inequality? One who purchases a stochastic cash flow for its minimum possible value can do no worse than that, and might do better. And one who purchases it for its maximum can do no better, and might do worse.

[1] Outcomes of zero probability are impossible outcomes, and do not affect the range of value.

As a corollary, if all the outcomes have the same deterministic value, the value of the stochastic cash flow must equal that value. For example, suppose that you value at 100 dollars both the receipt of 104 dollars one year from now and the receipt of 109 dollars two years from now. The stochastic cash flow is to flip a coin and to receive 104 dollars in one year if the coin lands heads, and to receive 109 dollars in two years if it lands tails. You should value the stochastic cash flow at 100 dollars. If the coin were flipped immediately after you paid the 100 dollars and you observed the outcome, you would think, "Fine. In effect, I bought this deterministic cash flow at its value." If your $v(x,t)$ function were compatible with the U.S. Treasury market, you could even cash in and regain your 100 dollars. But does it matter whether you will know the outcome immediately afterward? After all, no matter what will happen and no matter when you will learn of it, you will certainly receive something that right now you value at 100 dollars. That your $v(x,t)$ function may change (indeed, it almost certainly will change) makes no difference to the present.

True, one cannot hedge this stochastic cash flow so as at any moment to cash it in for 100 dollars (regardless of risk-free short-term interest). Some argue from this that the value of the flow should be less than 100 dollars. But the argument proves nothing because it proves too much. For it would undermine the $v(x,t)$ function itself. Just as little can one hedge the receipt of 104 dollars in one year so as at will to recover one's cost. So the hedging argument is no more germane to stochastic cash flows than it is to deterministic ones.

The implications of the stochastic-value inequality, $\min_i\{v(x_i,t_i)\} \le V[C] \le \max_i\{v(x_i,t_i)\}$, are revolutionary. Most likely, $V[C]$ should be not just *bounded* by the deterministic values, but even *determined* by them. In other words, the only property of outcome (x_i, t_i) that affects stochastic value is $v(x_i, t_i)$.[2] Therefore, the value of a stochastic cash flow should be invariant to any change in its outcomes that preserves their deterministic values. Furthermore, since nothing honors a minimum and a maximum better than a weighted average, it seems that stochastic value should be of the form:

$$V[C] = \sum_{i=1}^{n} q_i \cdot v(x_i, t_i)$$

Thus, to value a stochastic cash flow you should weight the deterministic values of its outcomes according to artificial, or risk-adjusted, probabilities. How to do this the author has shown in a full-length paper.[3] Suffice to say, a revolution is mounting against the ruling method of discounting expected cash flows at some risk-adjusted rate of return.[4] One should gauge the ruling method's right to his or her intellectual allegiance according to its eagerness and ability to respond to simple thought experiments like this one.

---

[2] Of course, the probabilities must not change. However, they are extrinsic to the outcomes.
[3] "The Valuation of Stochastic Cash Flows," CAS Forum (Reinsurance Discussion Papers, Spring 2003), 1-68.
[4] For detailed arguments against risk-adjusted discounting and its consort, capital allocation, see the author's paper "A Critique of Risk-Adjusted Discounting," *2001 ASTIN Colloquium*, www.casact.org/coneduc/reinsure/astin/2000/halliwell1.doc.

A Cash Flow Model for Forecasting Underwriting Investment Income

Louis B. Spore, ACAS, MAAA

A Cash Flow Model for Forecasting Underwriting Investment Income

Louis Spore, ACAS

Tower Group Companies

Abstract

The investment income received by a property-casualty company can be a prime component in its pricing and decision to write some lines of business that generate underwriting losses. In times of high interest rates it can enable the insurer to write during soft markets and to gain market share by taking on previously uninsurable risks.

Without the dynamic aspect of doing business the problem of investment income would reduce to watching a static amount of money, the surplus, accumulate in a savings or investment fund. The timing of the acquisition of new revenue, the uncertainty of the reserves and the payment of losses complicate the estimation of future investment income even if the amount of future written premium were known for a certainty. However a model that allows for a scenario testing of the random elements would be a useful tool in forecasting investment income.

A Cash Flow Model for Forecasting Underwriting Investment Income

Louis Spore

1. Introduction

Constructing a cash flow model involves the inevitable use of assumptions. In the following a deterministic model is created that uses certain simplifications. Among them is that the investment rate of return is known and does not vary from year to year. Another is that this rate is the same rate used to discount the reserves and a third is that the ultimate loss ratio has been accurately forecast. Since the principle impetus for the research done here is to construct a picture of insurance company financial position at future times under different operating conditions, these assumptions are simply consistent with the omniscience implied by a scenario analysis. To avoid too much simplicity the model is re-worked by allowing the rate of return to vary from year to year, although the rate used to discount the reserves is assumed to remain constant.

Some of the tricks for handling multiple sums and solving recurrence relations that are employed in the Appendices can be found in [KN].

2. The Cash Flow Algorithms

The first step in estimating investment income for an insurance company is to identify the cash that there is available to invest. Setting aside for the moment the cash in surplus and ongoing cash flow from current investments the source of new available cash has to be in the written business. We will assume that if the company stopped doing business immediately there would be enough cash and invested assets to pay the liabilities. Surplus would not be impacted and future investment income would come only from the surplus. If it does not stop business the increase in investment income will come from the operating cash flows from business written and not from a more clever investing strategy.

In line with these remarks we will start with a single policy written for at the beginning of the year net of re-insurance for an amount P. The cash invested from P will be P less the underwriting expense plus ceded commissions received. We will assume for the moment that premium is received, expenses are paid and, initially, reinsurance cessions are consummated and paid at the time the policy is first written. The errors induced by these timing assumptions can be refined later.

Let C_j be the loss fund at time t=j. If E is the underwriting expense and profit percentage, R is the ceding commission, W is the original written premium and c is the percentage ceded, then C_0, the initial cash fund , is $(1\text{-}c)W - EW + cWR$. In order for this fund to be positive the inequality $R \geq 1 - (1\text{-}E)/c$ should be true. Companies are shrewd enough negotiators to guarantee this will be a fact. In the exposition that follows the time j will refer to j years although it could be any time interval if the claim settlement pattern conforms to it. P=(1-c)W will be the retained premium that enters all of the formulas. [1]

There won't be a separate return on loss reserves and unearned premium. The entire reserve at the inception of the policy is LP and this will be reduced as payments are made. The return on year *n* will follow the recurrence:

$$I_n = iC_{n-1} - p_n LPr \qquad (2.1)$$

Here i is the rate of return, p_j is the percentage of ultimate paid in year j, L is the projected ultimate loss ratio and $r = (1+i)^{1/2}-1$ under the assumption that the amount paid in year j averages to the middle of the year. This simply states that the return during the year is from the cash fund at the beginning of the year less the interest lost from a payment made during the middle of the year.

The next task is to estimate the size of the fund that supports this policy at a future point in time, say *n*. The completion of this task involves first measuring the impact of taxes on the investments earned so far. We will define a tax variable, T_n at time n as the change in the discounted reserve less the amount paid during the year.

$$T_n = d_{n-1}R_{n-1} - d_n R_n - p_n LP \qquad (2.2)$$

Here d_n refers to a discount factor defined as

$$d_n = \frac{\sum_{k=n+1}^{N} p_k v^{k-n-1/2}}{\left(1-\sum_{k=1}^{n} p_k\right)} = \frac{LP\sum_{k=n+1}^{N} p_k v^{k-n-1/2}}{R_n} \qquad (2.3)$$

where N is the year of the last payment and $v = 1/(1+i)$

If S_j is the contribution made to the loss fund for some year $j \leq n$, then the loss fund at time n is

$$C_n = C_0 + \sum_{j=1}^{n} S_j - LP\sum_{j=1}^{n} p_j$$

$$\text{where } S_j = (1-t)\left[iC_{j-1} - rp_j LP\right] - tT_j \qquad (2.4)$$

$$\text{and } T_j = R_{j-1}d_{j-1} - R_j d_j - LPp_j$$

---

[1] Also, C_0 should be reduced by any excess premium ceded and the ultimate loss ratio modified by the impact that this might have on the retained losses.

for some tax rate t. If we call the first sum in (2.4) D_n and notice that $S_n = D_n - D_{n-1}$ we can use a standard trick for summing a series to get the following equation (see Appendix 1):

$$D_n = \sum_{j=1}^{n} S_j = C_0(u^n - 1) - (u-1)LP\sum_{j=1}^{n} u^{n-j}\sum_{k=1}^{j-1} p_k + LP[t - r(1-t)]\sum_{j=1}^{n} u^{n-j}p_j - t\sum_{j=1}^{n} \Delta R_j u^{n-j}$$

In this expression $u = 1+(1-t)i$ is the interest factor reduced by the tax rate and $\Delta R_j = d_jR_j - d_{j-1}R_{j-1}$. A few more algebraic manipulations (Appendices 1,2 and 3) give a simple equation for C_n.

$$C_n = (C_0 - LPd_0)u^n + R_nd_n \qquad (2.5)$$

If $n = N$, the year of the last payment, then $R_n = 0$. This also gives the break even loss ratio $L_b = C_0/d_0P$, which, in the absence of reinsurance, simplifies to $(1-E^*)/d_0$. E^* would be the underwriting expense ratio without a loading for the profit provision. This formula also gives a stronger expression for the minimum commission that the ceding company should demand from the re-insurer. At n=N it should be true that $C_0 = (1-c)W-EW+cRW \triangleleft_{\leftarrow}$

$d_0L(1-c)W$ or that $R \leftarrow 1-[(1-E)-d_0L(1-c)]/c$. Of course if L is too large the required R becomes too large and it becomes doubtful if there will be any quota share re-insurance.

3. Surplus

A related problem is determining if writing this policy will be profitable. This entails considering the return on the surplus, the investment return on underwriting and the underwriting profit. The fundamental equation to consider is:

$$Z_n = Z_0u^n + \sum_{k=1}^{n} S_k - LP\sum_{k=1}^{n} p_k + C_0$$

$$= Z_0u^n + C_n \qquad (3.1)$$

where Z_n is the surplus at the end of year n. The thing to notice about this formula is that the surplus increases because of two separate sources: the interest on the original surplus and the interest on the underwriting account. This can be a source of some confusion for an ongoing operation. After the first year Z_0 turns into Z_1, and for accounting purposes, this becomes the original surplus for a new year. It would not be correct to replace Z_0 with Z_1 in (3.1) because it would over-count the interest on the interest implicit in C_n. We must choose a zero year and then add the surplus contributions from future years by summing the C-contributions for each year from the zero year to the year n. For the purposes of calculating the income from the surplus Z_0

will not grow from underwriting until $n=N$, the year of the last payment and it is known for certain if there is a positive or negative contribution. Until that point is reached return on surplus for year n is just $(u-1)Z_0u^{n-1}$ and the return on the C_n contribution is calculated with (2.1). The surplus at the end of year N would be $Z_0u^N+C_N$.

The exception to this is the issue of capital infusion. For example, capital can be supplied by a surplus note, private investment or a stock issue. The timing can also be critical. If the *n-th* contribution , CC_n, is made at time T_n with initial surplus of Z_0 in year 1, the surplus for year n will be $Z_n = uZ_{n-1} + u^{1-T_n}CC_n$. Here $1 \trianglelefteq T_n \trianglelefteq 0$ is the fraction of the year that represents when the contribution is received by the company. Thus a contribution effective at June 30th of the year would make $T_n = ½$. This implies a general expression to use for computing surplus return: $Z_n = Z_0u^n + \sum_{k=1}^{n} u^{n+1-k-T_k}CC_k$. The income from surplus for year n would then be $(u-1)Z_{n-1} + (u^{1-T_n} - 1)CC_n$. Appendix 4 shows the from the equation for Z_n takes under the assumptions that the rate of return is different from year to year but that the rate for discounting reserves remains constant.

It almost goes without saying that if the capital contribution is from a surplus note, the calculated return should be reduced by the amount of interest payable on the note at the end of the year.

Exhibit 1 illustrates the flow into future years of a single years written premium. The last line uses the formula to calculate the cash flow fund as a check on the cash flow lines above.

4. Calendar Year Investment Income

This analysis began with the hypothesis that P is written premium and that the loss experience is policy year loss experience. At the end of a calendar year only half of the written premium is earned and only half of the losses have been incurred. The tax treatment of the reserves is on the reserves for losses that have been incurred. The derivation began with a single policy written at the beginning of the year. There isn't anything in the derivation that prevents us from regarding the premium earned during the year as coming from a single policy written at the beginning of the year for the amount of the earned premium at its end. The reserves will be accident year reserves, the p_j will be accident year settlement patterns and the tax treatment will be on accident year reserves. Throughout the assumption will be that the accident year outstanding and IBNR reserves for all years is accurate and therefore the accident year ultimate losses are equal to the calendar year incurred losses.

Assuming that premium writings stay constant from year to year, the calendar year investment income at year n would be the sum of equation (2.1) from one to n.

$$\sum_{k=1}^{n} I_k = (1-t)\sum_{k=1}^{n}\left[iC_{k-1} - rp_k LP\right]$$

$$= \sum_{k=1}^{n} S_k + t\sum_{k=1}^{n}\Delta R_k - tLP\sum_{k=1}^{n} p_k$$

$$= (C_0 - LPd_0)(u^n - 1) + LP(1-t)\sum_{k=1}^{n} p_k - LP(1-t)d_0$$

An alternate derivation of this formula and confirmation of its correctness would be to derive it directly from equation (2.5)

$$\sum_{k=1}^{n} I_k = \sum_{k=1}^{n} i(1-t)C_{k-1} - rLP(1-t)\sum_{k=1}^{n} p_k$$

$$= i(1-t)\sum_{k=1}^{n}(C_0 - LPd_0)u^{k-1} - rLP(1-t)\sum_{k=1}^{n} p_k + i(1-t)\sum_{k=1}^{n} R_{k-1}d_{k-1}$$

$$= \frac{(u-1)(C_0 - LPd_0)(u^n - 1)}{(u-1)} - rLP(1-t)\sum_{k=1}^{n} p_k + LPi(1-t)\sum_{k=1}^{n}\sum_{j=k}^{n} v^{j-k+1/2} p_j$$

$$= (C_0 - LPd_0)(u^n - 1) - rLP(1-t)\sum_{k=1}^{n} p_k + LPi(1-t)\sum_{j=1}^{n}\left(\sum_{k=1}^{j} v^{-k}\right)v^{j+1/2} p_j$$

$$= (C_0 - LPd_0)(u^n - 1) - rLP(1-t)\sum_{k=1}^{n} p_k - v^{-1}LP(1-t)\sum_{j=1}^{n}(1 - v^{-j})v^{j+1/2} p_j$$

$$= (C_0 - LPd_0)(u^n - 1) - rLP(1-t)\sum_{k=1}^{n} p_k - LP(1-t)d_0 + v^{-1/2}LP(1-t)\sum_{k=1}^{n} p_k$$

$$= (C_0 - LPd_0)(u^n - 1) + LP(1-t)\sum_{k=1}^{n} p_k - LP(1-t)d_0$$

This is all very interesting but it is rarely true that premium is written at the same level each year or that the ultimate loss ratios are the same from year to year. The advantage of simplified formulas is to provide insight into the process of how investment income is related to the underwriting process. There is still some practical use for formula (2.5) in constructing a pro-forma to predict future investment income under different written premium and re-insurance scenarios. We can find the cash fund in year *n* to support different levels of premium, reinsurance, interest rate, ultimate loss ratio and underwriting expense assumptions for all years prior to *n*. We could then use formula (2.1) to find the investment contribution each policy year makes to the current calendar year.

5. Refinements

A realistic scenario often includes the cession of re-insurance earned premium on a quarterly basis with the payment of commission based solely on the amounts ceded. This can result in an increase in the amounts of interest on the delay of premium payments and a decrease of the return on the commission not yet received. There can be other things that affect surplus such as capital infusions or surplus notes. If the capital infusion comes from a surplus note the interest on the note would be a deduction to the rate of return.

Accounting for these things can be simple enough in a spreadsheet format. We can regard the additional interest or additional outgo as adjustments to surplus. For example if written premium is ceded quarterly on an earned basis the interest on one dollar of premium written at the beginning of the year at the end of the year would be equal to:

$$i + 0.5 - \sum_{k=1}^{4} (1+i)^{1-k/4} \frac{2k-1}{32}$$

This amount would be added to surplus and form part of the next year's initial surplus.

In practice, the IRS rules for taxes are more complex than the simple allowance for the change in discounted reserves and the amounts paid shown in the model. For example there is a credit for 80% of the change in the unearned premium reserve and 70% of the dividends received are deductible. However it isn't the purpose of the model to follow the impact of taxes on income but only the impact that taxes have on investment income that is derived directly from underwriting. The choice of the rate of return should be based on the historical patterns of net return and thus would incorporate the special treatment given to dividends or realized capital gains. The details of the tax treatment for P&C companies can be found in [AG].

We can also generalize the expression for the cash fund, equation (2.5), by assuming that rates of return differ each year (and is = i_n for year n). A separate (constant) rate i is assumed for discounting reserves. Appendices 5-7 show the derivation of the more general equation to be

$$(C_0 - LPd_0)\prod_{m=1}^{n} u_m + \alpha_n R_n d_n + \beta_n (1+r)LP - LP(1-t)\gamma_n (1+r_n)p_n \qquad (5.1)$$

where u_m is the tax adjusted interest factor for year m and $\alpha_n = ivt + \alpha_{n-1} v u_n$

$\beta_n = (\alpha_n - t)p_n + \beta_{n-1} u_n$

$\gamma_n = \dfrac{(1+r_{n-1})p_{n-1} u_n \gamma_{n-1}}{(1+r_n)p_n} + 1$ where $\alpha_0 = 1$, $\beta_0 = 0$, $\gamma_0 = 0$ and $r_n = \sqrt{1+i_n} - 1$

This looks dauntingly complicated but the practical computation doesn't present any problems.[2] It will give us another equation for calculation of the breakeven loss ratio however, because it is not generally true that $\updownarrow_N = 0 = \bullet_N$. The new formula becomes

$$L_b = \frac{C_0}{P\left\{d_0 - \prod_{m=1}^{N} u_n^{-1}\left[\beta_N(1+r) - (1-t)\gamma_N p_N(1+r_N)\right]\right\}} \tag{5.2}$$

Whether this is an increase or decrease to the breakeven loss ratio calculated earlier depends on the average size of the rates of return. If they are small from year to year and less than the discount rate, $\updownarrow_N$ will be negative and $\bullet_N$ positive. The new breakeven ratio will be smaller. Large rates of return will make $\updownarrow_N$ positive and $\bullet_N$ close to zero. The breakeven ratio will be larger. An illustration of the calculation of $\mapsto$ $\updownarrow$ and $\bullet$ for small and large rates of return is shown in Exhibit 2.

Although Appendices 5-7 follow the logic of Appendices 1-3, Appendix 7 differs from Appendix 3 by proving a formula for $-t\lhd\lambda R_j$ by induction. Whenever the *u*'s are the same from year to year, and use a rate equal to the discount rate, it isn't to hard to see that the alphas are all identically equal to 1 and the beta and gamma factors combine to equal zero.

6. Summary

The cash flow equations for a deterministic model of future investment income were derived from a consideration of an initial cash fund that is derived from the underwriting process. The recurrence relations were based on the return from this fund and the tax implications of the change in the discounted reserves and the amounts paid. The term "reserves" as defined in the model is the premium times the estimated ultimate loss and loss adjustment expense ratio. At the beginning of the year this would include "policy reserves" or reserves for claims not yet incurred and the unearned premium reserves.

The analysis begins by following the cash flow to support a policy written at the beginning of the year by establishing a fund to pay claims from the premium written and the offsets to it represented by the re-insurance agreements and the underwriting expenses. The calculation of the investment return follows by applying the expected rate of return to the formula for the cash fund at the prior year.

The model is first generalized by assuming first that the premium earned during the year is equivalent to a single policy written at the beginning of the year, and then by assuming different rates of return to be applicable for different calendar years.

---

2 Appendix 8 gives closed form solutions to the alpha , beta and gamma recurrences

References:

[KN] Graham, Ronald L; Knuth, Donald E and Patashnik, Oren, *Concrete Mathematics,* Addison-Wesley, 2^{nd} Edition 1998.

[AG] M. Almagro, T.L. Ghezzi, , "Federal Income Taxes-Provisions Affecting Property/Casualty Insurers", PCAS LXXV, 1988

[F] Ferrari, J.Robert, "The Relationship of Underwriting, Investment, Leverage and Exposure to Total Return on Owners' Equity", PCAS LV, 1968

[K] Kellison, Stephen G., *The Theory of Interest,* Irwin, 2^{nd} Edition 1991.

Exhibit 1 - Sheet 1

The Yearly Rate of Return and the Discount Rate Are All the Same

Return (discount) rate = 4.0%

	r = 6-month rate =	1.98%
	u = 1+i(1-t) =	1.0240
	d_0 =	0.8492
	Breakeven Ult. L/R	68.6%

	Year	1	2	3	4	5	6	7	8	9	10	11
(1)	Prior Surplus	2,213,780	4,172,496	3,984,224	3,738,464	3,488,869	3,268,365	3,093,634	2,969,421	2,892,791	2,856,773	2,853,048
(2)	LOB DWP (W)	7,000,000										
(3)	Premium for Excess (x)	146,661										
(4)	Ceded %	50.0%										
(5)	Net DWP [(W-x)(1-c)=P]	3,426,669										
(6)	Expense ratio (E)	20.0%										
(7)	Expenses (WE)	1,400,000										
(8)	Ultimate Loss ratio (L)	68.6%										
(9)	Commission rate	40.0%										
(10)	Commissions (cR)(W-x)	1,370,668										
(11)	**Beginning Loss Reserve**	**2,352,139**	**2,183,826**	**1,871,118**	**1,510,276**	**1,156,188**	**841,310**	**580,881**	**377,782**	**227,176**	**120,382**	**47,674**
(12)	% paid during year	7.2%	13.3%	15.3%	15.1%	13.4%	11.1%	8.6%	6.4%	4.5%	3.1%	2.0%
(13)	Paid on reserve (LP_R)	168,312	312,708	360,842	354,088	314,878	260,429	203,099	150,606	106,794	72,708	47,674
(14)	Interest on paid (LPr_R)	3,333	6,193	7,146	7,012	6,236	5,158	4,022	2,983	2,115	1,440	944
(15)	**Ending Loss Reserve (R_n)**	**2,183,826**	**1,871,118**	**1,510,276**	**1,156,188**	**841,310**	**580,881**	**377,782**	**227,176**	**120,382**	**47,674**	**0**
(16)	Reserve Discount factor (d_n)	0.8726	0.8887	0.9014	0.9123	0.9222	0.9319	0.9419	0.9530	0.9656	0.9806	1.0000
(17)	Discounted Reserve ($R_n d_n$)	1,905,585	1,662,907	1,361,435	1,054,792	775,870	541,319	355,850	216,495	116,246	46,748	0
(18)	Change in Disc'td Reserve	91,752	242,678	301,472	306,643	278,922	234,551	185,468	139,355	100,249	69,498	46,748
(19)	Tax Effect ($-tT_n$)	30,624	28,012	23,748	18,978	14,382	10,351	7,052	4,501	2,618	1,284	370
###	*Cash Fund (beginning)*	*1,997,337*	*1,905,585*	*1,662,907*	*1,361,435*	*1,054,792*	*775,870*	*541,319*	*355,850*	*216,495*	*116,246*	*46,748*
(21)	**Interest on Reserve: ($IC_{n-1} - rp_nLP$)**	76,560	70,031	59,370	47,445	35,956	25,877	17,631	11,251	6,545	3,210	926
(22)	**$(1-t)(IC_{n-1} - rp_nLP)-tT_n$**	**76,560**	**70,031**	**59,370**	**47,445**	**35,956**	**25,877**	**17,631**	**11,251**	**6,545**	**3,210**	**926**
###	*Cash Fund (end)*	*1,905,585*	*1,662,907*	*1,361,435*	*1,054,792*	*775,870*	*541,319*	*355,850*	*216,495*	*116,246*	*46,748*	*0*
(24)												
(25)	TAX Rate	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%
(26)												
(27)	Return on Initial Surplus**	53,131	54,406	55,712	57,049	58,418	59,820	61,256	62,726	64,231	65,773	67,351
(28)	**Net addition to surplus: (22)+(27)-(13)**	**-38,622**	**-188,272**	**-245,761**	**-249,594**	**-220,504**	**-174,731**	**-124,213**	**-76,629**	**-36,018**	**-3,725**	**20,603**
(29)	Ending Surplus: (28) + (1)*	4,172,496	3,984,224	3,738,464	3,488,869	3,268,365	3,093,634	2,969,421	2,892,791	2,856,773	2,853,048	2,873,651
(30)												
(31)	$C_n = (C_0 - LPd_0)u^n + R_nd_n$	1,905,585	1,662,907	1,361,435	1,054,792	775,870	541,319	355,850	216,495	116,246	46,748	0
(32)												
(33)	Investment Income (before taxes)	76,560	70,031	59,370	47,445	35,956	25,877	17,631	11,251	6,545	3,210	926
(34)	Cumulative Investment Income	76,560	146,591	205,961	253,406	289,362	315,239	332,870	344,121	350,666	353,876	354,802
(35)												
(36)	Z_0u^n	2,266,911	2,321,317	2,377,029	2,434,077	2,492,495	2,552,315	2,613,571	2,676,296	2,740,527	2,806,300	2,873,651
(37)	$Z_0u^n+C_n$	4,172,496	3,984,224	3,738,464	3,488,869	3,268,365	3,093,634	2,969,421	2,892,791	2,856,773	2,853,048	2,873,651

* This is equal to (28) + (1) + (20) for the first year

** $(u-1)Z_0$+ (u-1) x sum of the entries for prior years

Exhibit 1 - Sheet 2

The Yearly Rate of Return Differs from Year to Year

Discount rate = 4.0%

	Year	1	2	3	4	5	6	7	8	9	10	11
(1)	Prior Surplus	2,213,780	4,189,940	4,035,363	3,830,738	3,631,308	3,471,192	3,308,117	3,217,122	3,185,805	3,207,666	3,275,163
(2)	LOB DWP (W)	7,000,000										
(3)	Premium for Excess (x)	146,661										
(4)	Ceded %	50.0%		r = 6-month rate =		1.98%						
(5)	Net DWP [(W-x)(1-c)=P]	3,426,669		d_0 =		0.8492						
(6)	Expense ratio (E)	20.0%		β_N =		0.844	0.860 =$(1+r)\beta_N$					
(7)	Expenses (WE)	1,400,000		γ_N =		63.133	0.796 =$(1-t)(1+r_N)p_N\gamma_N$					
(8)	Ultimate Loss ratio (L)	72.2%		$1/\Pi u_j$ =		0.657						
(9)	Commission rate	40.0%		Breakeven Ult. L/R		72.2%						
(10)	Commissions (cR)(W-x)	1,370,668										
(11)	**Beginning Loss Reserve**	**2,475,100**	**2,297,989**	**1,968,934**	**1,589,228**	**1,216,629**	**885,291**	**611,248**	**397,531**	**239,052**	**126,675**	**50,166**
(12)	% paid during year	7.2%	13.3%	15.3%	15.1%	13.4%	11.1%	8.6%	6.4%	4.5%	3.1%	2.0%
(13)	Paid on reserve (LPp_n)	177,111	329,056	379,706	372,598	331,339	274,043	213,716	158,480	112,377	76,509	50,166
(14)	Interest on paid ($LPrp_n$)	3,507	6,517	7,520	7,379	6,562	5,427	4,232	3,139	2,225	1,515	993
(15)	**Ending Loss Reserve (R_n)**	**2,297,989**	**1,968,934**	**1,589,228**	**1,216,629**	**885,291**	**611,248**	**397,531**	**239,052**	**126,675**	**50,166**	**0**
(16)	Reserve Discount factor (d_n)	0.8726	0.8887	0.9014	0.9123	0.9222	0.9319	0.9419	0.9530	0.9656	0.9806	1.0000
(17)	Discounted Reserve (R_nd_n)	2,005,202	1,749,838	1,432,606	1,109,933	816,430	569,617	374,453	227,813	122,323	49,192	0
(18)	Change in Disc'td Reserve	96,549	255,364	317,232	322,673	293,503	246,813	195,164	146,640	105,490	73,131	49,192
(19)	Tax Effect ($-tT_n$)	32,225	29,477	24,990	19,970	15,134	10,892	7,421	4,736	2,755	1,351	390
(20)	Cash Fund (beginning)	1,997,337	1,909,747	1,673,082	1,376,328	1,073,813	798,610	555,357	365,263	222,723	120,135	48,694
(21)	Rate of Return (i_n)	5.0%	6.0%	6.5%	7.0%	7.5%	5.0%	6.0%	6.5%	7.0%	7.5%	7.5%
(22)	u_n	1.030	1.036	1.039	1.042	1.045	1.030	1.036	1.039	1.042	1.045	1.045
(23)	α_n	1.006	1.017	1.032	1.049	1.069	1.075	1.086	1.100	1.118	1.138	1.159
(24)	β_n	0.043	0.127	0.229	0.336	0.441	0.529	0.607	0.676	0.737	0.793	0.844
(25)	γ_n	1.000	1.555	2.397	3.539	5.149	7.489	10.901	16.238	24.806	38.986	63.133
(26)	**Interest on Reserve: ($i_nC_{n-1} - r_np_nLP$)**	95,493	104,857	96,604	83,523	68,335	33,163	27,003	18,673	11,724	6,193	1,805
(27)	**$(1-t)(i_nC_{n-1} - r_np_nLP)-tT_n$**	89,521	92,391	82,952	70,084	56,135	30,790	23,623	15,939	9,789	5,067	1,473
(28)	Cash Fund (end)	1,909,747	1,673,082	1,376,328	1,073,813	798,610	555,357	365,263	222,723	120,135	48,694	0
(29)	TAX Rate	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%
(30)	Return on Initial Surplus**	66,413	82,087	92,129	103,085	115,087	80,177	99,099	111,223	124,449	138,939	145,191
(31)	**Net addition to surplus: (27)+(30)-(13)**	-21,177	-154,578	-204,625	-199,430	-160,116	-163,076	-90,994	-31,318	21,862	67,497	96,498
(32)	Ending Surplus: (31) + (1)*	4,189,940	4,035,363	3,830,738	3,631,308	3,471,192	3,308,117	3,217,122	3,185,805	3,207,666	3,275,163	3,371,661
(33)												
(34)	**$C_n = (C_0 - LPd_0)\Pi u_j + \alpha_nR_nd_n$**											
(35)	**$+ \beta_n(1+r)LP - (1-t)LP\gamma_n(1+r_n)p_n$**	1,909,747	1,673,082	1,376,328	1,073,813	798,610	555,357	365,263	222,723	120,135	48,694	0
(36)	Investment Income (before taxes)	95,493	104,857	96,604	83,523	68,335	33,163	27,003	18,673	11,724	6,193	1,805
(37)	Cumulative Investment Income	95,493	200,350	296,954	380,477	448,812	481,975	508,978	527,651	539,375	545,568	547,373
(38)												
(39)	**$Z_0\Pi u_j$**	2,280,194	2,362,281	2,454,410	2,557,495	2,672,582	2,752,760	2,851,859	2,963,082	3,087,531	3,226,470	3,371,661
(40)	**$Z_0\Pi u_j+C_n$**	4,189,940	4,035,363	3,830,738	3,631,308	3,471,192	3,308,117	3,217,122	3,185,805	3,207,666	3,275,163	3,371,661

* This is equal to (20) + (1) + (31) for the first year

** $(\Pi u_j-1)Z_0$ - sum of prior year entries

Exhibit 2

Calculation of the Alpha and Beta Factors for Varying Interest Rates in the General Model

Tax Rate	40.0%	1+r =	1.024695
Disc't Rate	5.0%	β_N =	0.69380
$1/\Pi u_n$ =	0.793	γ_N =	195.97931

$(1+r)\beta_N/\Pi u_n$ = 0.56392

$(1-t)p_N(1+r_N)\gamma_N/\Pi u_n$ = 0.56796

$(1+r)\beta_N/\Pi u_n$ - $(1-t)p_N(1+r_N)\gamma_N/\Pi u_n$ = -0.00404

Calculation of Alpha and Beta for Low Rates of Return

n	1	2	3	4	5	6	7	8	9	10	11	12
i_n	4.0%	3.5%	3.7%	3.0%	2.5%	3.6%	3.9%	4.0%	2.5%	3.3%	2.0%	3.0%
u_n	1.024	1.021	1.022	1.018	1.015	1.022	1.023	1.024	1.015	1.020	1.012	1.018
r_n	2.0%	1.7%	1.8%	1.5%	1.2%	1.8%	1.9%	2.0%	1.2%	1.6%	1.0%	1.5%
p_n	23.6%	21.0%	16.1%	12.0%	8.7%	6.2%	4.3%	3.0%	2.1%	1.4%	1.0%	0.6%
α_n	0.994	0.991	0.993	0.989	0.986	0.992	0.994	0.994	0.986	0.990	0.983	0.989
β_n	0.14025	0.26740	0.36874	0.44600	0.50365	0.55123	0.58966	0.62164	0.64327	0.66427	0.67807	0.69380
γ_n	1.00000	2.15018	3.86408	6.29552	9.83520	15.02406	23.13747	34.94331	52.03719	80.29246	115.48100	195.97931

Tax Rate	40.0%	1+r =	1.024695
Disc't Rate	5.0%	β_N =	0.92492
$1/\Pi u_n$ =	0.565	γ_N =	1.13404

$(1+r)\beta_N/\Pi u_n$ = 0.53561

$(1-t)p_N(1+r_N)\gamma_N/\Pi u_n$ = 0.00239

$(1+r)\beta_N/\Pi u_n$ - $(1-t)p_N(1+r_N)\gamma_N/\Pi u_n$ = 0.53322

Calculation of Alpha and Beta for High Rates of Return

n	1	2	3	4	5	6	7	8	9	10	11	12
i_n	10.0%	8.8%	9.3%	7.5%	6.3%	9.0%	9.8%	10.0%	6.3%	8.3%	5.0%	7.5%
u_n	1.060	1.053	1.056	1.045	1.038	1.054	1.059	1.060	1.038	1.050	1.030	1.045
r_n	4.9%	4.3%	4.5%	3.7%	3.1%	4.4%	4.8%	4.9%	3.1%	4.0%	2.5%	3.7%
p_n	23.6%	21.0%	16.1%	12.0%	8.7%	6.2%	4.3%	3.0%	2.1%	1.4%	1.0%	0.6%
α_n	1.029	1.021	1.024	1.014	1.007	1.023	1.027	1.029	1.007	1.019	1.000	1.014
β_n	0.14834	0.28663	0.40305	0.49490	0.56628	0.63548	0.69962	0.76045	0.80172	0.85007	0.88157	0.92492
γ_n	1.00000	1.09869	1.13237	1.11453	1.09643	1.13736	1.15954	1.16608	1.10526	1.13598	1.08028	1.13404

Exhibit 3

Calculation of the Reserve Discount Factor

Interest rate = 4%

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Year	Discount Factor	Incremt'l Payments	Cumulative Payments	Complement of Cumul.	Product	Cumul Product	Comp. Of Product	Reserve Discount Factor (d_n)
1	0.9806	7%	7%	93%	0.0702	0.0702	0.7790	0.8726
2	0.9429	13%	20%	80%	0.1254	0.1955	0.6536	0.8887
3	0.9066	15%	36%	64%	0.1391	0.3346	0.5146	0.9014
4	0.8717	15%	51%	49%	0.1312	0.4658	0.3833	0.9123
5	0.8382	13%	64%	36%	0.1122	0.5780	0.2711	0.9222
6	0.8060	11%	75%	25%	0.0892	0.6673	0.1819	0.9319
7	0.7750	9%	84%	16%	0.0669	0.7342	0.1150	0.9419
8	0.7452	6%	90%	10%	0.0477	0.7819	0.0673	0.9530
9	0.7165	5%	95%	5%	0.0325	0.8144	0.0347	0.9656
10	0.6889	3%	98%	2%	0.0213	0.8357	0.0134	0.9806
11	0.6624	2%	100%	0%	0.0134	0.8492	0.0000	1.0000
		100%			0.8492	$=d_0$		

Notes:

Col (2): 1.04^[Col (1) - 1/2]
Col (3): from Exhibit A
Col (4): Cumulative of Col (3)
Col (5): 1- Col (4)

Col (6): Col (2) x Col (3)
Col (7): Cumulative of Col (6)
Col (8): Sum of Col (7) - Col (7)
Col (9): 1.04^[Col (1)] x Col (8) / Col (5)

Exhibit 4

Mathematical Formulation of Exhibit 3

For: i = interest rate
p_j = % of ultimate paid in year j

$$Col\ (1) : k$$

$$Col\ (2) : (1+i)^{-k+1/2}$$

$$Col\ (3) : p_k$$

$$Col\ (4) : \sum_{j=1}^{k} p_j$$

$$Col\ (5) : 1 - Col\ (4) = \sum_{j=k+1}^{n} p_j$$

$$Col\ (6) : (1+i)^{-k+1/2}\, p_k$$

$$Col\ (7) : \sum_{j=1}^{k} (1+i)^{-j+1/2}\, p_j$$

$$Col\ (8) : \sum_{j=k+1}^{n} (1+i)^{-j+1/2}\, p_j$$

$$Col\ (9) : \frac{\sum_{j=k+1}^{n} (1+i)^{-j+k+1/2}\, p_j}{\sum_{j=k+1}^{n} p_j}$$

APPENDIX 1

$C_n = C_0 + \sum_{j=1}^{n} S_j - LP\sum_{j=1}^{n} p_j$ Now define $T_j = d_{j-1}R_{j-1} - d_jR_j - LPp_j$

where $R_j = LP\sum_{k=j+1}^{N} p_k$

where $S_j = (1-t)[iC_{j-1} - rp_jLP] - tT_j$ Letting $D_j = \sum_{k=1}^{j} S_k$ we get

$D_j - D_{j-1} = (1-t)[iC_{j-1} - rp_jLP] - tT_j$

$= (1-t)\left[i\left(C_0 + D_{j-1} - LP\sum_{k=1}^{j-1} p_k\right) - rp_jLP\right] + tLPp_j - t\Delta R_j$ (where $\Delta R_j = d_{j-1}R_{j-1} - d_jR_j$)

$= (u-1)D_{j-1} + \lambda_j$ if we let λ_j represent all of the terms that do not involve D and $u = 1+(1-t)i$

Hence if we multiply both sides of this equation by u^{-j} and sum we get

$\sum_{j=1}^{n}\left(D_ju^{-j} - D_{j-1}u^{-(j-1)}\right) = D_nu^{-n} = \sum_{j=1}^{n}\lambda_ju^{-j}$ because $D_0 = 0$ by its definition. Thus

$D_n = \sum_{j=1}^{n} S_j = \sum_{j=1}^{n}\lambda_ju^{n-j} = S_1^* - S_2^* + S_3^* + S_4^*$

$S_1^* = (u-1)C_0\sum_{j=1}^{n} u^{n-j} = C_0(u^n - 1)$

$S_2^* = (u-1)LP\sum_{j=1}^{n} u^{n-j}\sum_{k=1}^{j-1} p_k = LP\left(\sum_{k=1}^{n-1} p_k\left(u^{n-k} - 1\right)\right) = LPu^nA_{n-1} - LP\sum_{k=1}^{n-1} p_k$ {See Appendix 2}

for $A_n = \sum_{k=1}^{n} p_ku^{-k}$

$S_3^* = LP[t - r(1-t)]\sum_{j=1}^{n} u^{n-j}p_j = LP[t - r(1-t)]u^nA_n$

$S_4^* = -t\sum \Delta R_ju^{n-j} = (1-t)LPu^n(1+r)A_n + R_nd_n - LPd_0u^n$ {See Appendix 3}

Adding together these simplified sums we get

$$\sum_{j=1}^{n} S_j = C_0(u^n - 1) - \left[LPu^n A_{n-1} - LP\sum_{k=1}^{n-1} p_k \right] + \left\{ LP\left[t - r(1-t)\right]u^n A_n \right\}$$

$$+ (1 - t)LPu^n\left[(1 + r)A_n - d_0\right] + R_n d_n - tLPd_0 u^n$$

$$= C_0(u^n - 1) + LP\left[t - r(1 - t) + (1 - t)(1 + r)\right]u^n A_n - LPu^n A_{n-1} + LP\sum_{k=1}^{n-1} p_k - d_0 LPu^n + R_n d_n$$

$$= C_0(u^n - 1) + LPu^n\left(A_n - A_{n-1}\right) + LP\sum_{k=1}^{n-1} p_k - d_0 LPu^n + R_n d_n$$

[since $t - r(1 - t) + (1 - t)(1 + r) = 1$]

$$= C_0(u^n - 1) + LPu^n(p_n u^{-n}) + LP\sum_{k=1}^{n-1} p_k - d_0 LPu^n + R_n d_n$$

$$= C_0(u^n - 1) + LP\sum_{k=1}^{n} p_k - d_0 LPu^n + R_n d_n$$

Adding this to the expression for C_n we get

$$C_n = C_0 u^n - d_0 LPu^n + R_n d_n$$

APPENDIX 2

Simplification of the S_2^* term in the expression for the sum of the surplus contributions

$$(u-1)LP\sum_{j=1}^{n} u^{n-j}\sum_{k=1}^{j-1} p_k$$

$$= (u-1)LP\sum_{j=2}^{n} u^{n-j}\sum_{k=1}^{j-1} p_k$$

$$= (u-1)LP\sum_{j=1}^{n-1} u^{n-j-1}\sum_{k=1}^{j} p_k = (u-1)LP\sum_{k=1}^{n-1} p_k \sum_{j=k}^{n-1} u^{n-j-1}$$

$$= (u-1)u^{n-1}LP\sum_{k=1}^{n-1} p_k u^{-k}\sum_{j=k}^{n-1} u^{k-j}$$

$$= (u-1)u^{n-1}LP\sum_{k=1}^{n-1} p_k u^{-k}\left(\frac{1-u^{-(n-k)}}{1-u^{-1}}\right) = LP\sum_{k=1}^{n-1} p_k u^{-k}\left(u^{n}-u^{k}\right)$$

$$= LPu^{n}A_{n-1} - LP\sum_{k=1}^{n-1} p_k \quad \text{where } A_n = \sum_{k=1}^{n} p_k u^{-k}$$

APPENDIX 3

Simplification of the S_4^* term in the expression for the sum of the surplus contributions

$$t\sum_{j=1}^{n} R_{j-1} d_{j-1} u^{n-j}$$

$$= tLP\sum_{j=1}^{n} u^{n-j} \sum_{k=j}^{n} p_k v^{k-j+1/2} + tLP\sum_{j=1}^{n} u^{n-j} \sum_{k=n+1}^{N} p_k v^{k-j+1/2}$$

$$= tu^n LP\sum_{k=1}^{n} p_k v^{k+1/2} \sum_{j=1}^{k} (uv)^{-j} + tu^n LP \sum_{k=n+1}^{N} p_k v^{k+1/2} \sum_{j=1}^{n} (uv)^{-j}$$

$$= tLPu^{n-1} v^{-1} \sum_{k=1}^{n} p_k \left[\frac{1-(uv)^{-k}}{1-(uv)^{-1}}\right] v^{k+1/2} + tLPu^{n-1} v^{-1} \sum_{k=n+1}^{N} p_k \left[\frac{1-(uv)^{-n}}{1-(uv)^{-1}}\right] v^{k+1/2}$$

$$= -\frac{(1+i)}{i} LPu^n \left[\sum_{k=1}^{n} p_k v^{k+1/2} - v^{1/2} \sum_{k=1}^{n} p_k u^{-k}\right] - \frac{(1+i)}{i} LPu^n v^{n+1} \sum_{k=n+1}^{N} p_k v^{k-n-1/2} \left[1-(uv)^{-n}\right]$$

$$= \frac{LPu^n}{i} \left[v^{-1/2} A_n - \sum_{j=1}^{n} p_j v^{j-1/2}\right] + \frac{R_n d_n}{i} \left[1-(uv)^n\right]$$

$$= \frac{LPu^n}{i} \left[(1+r)A_n - \left(\sum_{j=1}^{N} p_j v^{j-1/2} - \sum_{j=n+1}^{N} p_j v^{j-1/2}\right)\right] + \frac{R_n d_n}{i} \left[1-(uv)^n\right]$$

$$= \frac{LPu^n}{i} \left[(1+r)A_n - \left(d_0 - v^n \sum_{j=n+1}^{N} p_j v^{j-n-1/2}\right)\right] + \frac{R_n d_n}{i} \left[1-(uv)^n\right]$$

$$= \frac{LPu^n}{i} \left[(1+r)A_n - d_0\right] + \frac{(uv)^n R_n d_n}{i} + \frac{R_n d_n}{i} \left[1-(uv)^n\right]$$

$$= \frac{LPu^n}{i} \left[(1+r)A_n - d_0\right] + \frac{R_n d_n}{i}$$

But

$$t\sum_{j=1}^{n} R_j d_j u^{n-j}$$

$$= t\sum_{j=2}^{n+1} R_{j-1} d_{j-1} u^{n-j+1}$$

$$= ut\sum_{j=1}^{n} R_{j-1} d_{j-1} u^{n-j} + tR_n d_n - tR_0 d_0 u^n$$

$$= u\left\{\frac{LPu^n}{i}\left[(1+r)A_n - d_0\right]\right\} + \frac{uR_n d_n}{i} + tR_n d_n - tLPd_0 u^n$$

since $R_0 = LP$

Hence

$$-t\sum_{j=1}^{n} \Delta R_j u^{n-j}$$

$$= (u-1)\left\{\frac{LPu^n}{i}\left[(1+r)A_n - d_0\right]\right\} + \frac{(u-1)R_n d_n}{i} + tR_n d_n - tLPd_0 u^n$$

$$= (1-t)LPu^n\left[(1+r)A_n - d_0\right] + R_n d_n - tLPd_0 u^n$$

$$= (1-t)LPu^n (1+r)A_n + R_n d_n - LPd_0 u^n$$

APPENDIX 4

The Derivation of "Interest" Surplus When Interest Rates Vary From Year to Year

Let $Z_k = u_k Z_{k-1} + u_k^{1-T_k} CC_k$

where $u_k = 1 + (1-t)i_k$ is the interest factor for year k

and T_k is the timing of the capital infusion for year k

$$Z_k \prod_{j=1}^{k} u_j^{-1} - Z_{k-1} \prod_{j=1}^{k} u_j^{-1} u_k = u_k^{1-T_k} \prod_{j=1}^{k} u_j^{-1} CC_k$$

$$\Rightarrow Z_n \prod_{j=1}^{n} u_j^{-1} - Z_0 = \sum_{k=1}^{n} u_k^{1-T_k} \left[\prod_{j=1}^{k} u_j^{-1} \right] CC_k$$

$$\Rightarrow Z_n = Z_0 \prod_{j=1}^{n} u_j + \sum_{k=1}^{n} u_k^{1-T_k} \left[\prod_{j=1}^{k} u_j^{-1} \right] \left[\prod_{j=1}^{n} u_j \right] CC_k$$

$$= Z_0 \prod_{j=1}^{n} u_j + \sum_{k=1}^{n} u_k^{1-T_k} \left[\prod_{j=k+1}^{n} u_j \right] CC_k$$

where the product in the second term is interpreted as $= 1$ if $k + 1 > n$

APPENDIX 5

$C_n = C_0 + \sum_{j=1}^{n} S_j - LP\sum_{j=1}^{n} p_j$ Now define $T_j = d_{j-1}R_{j-1} - d_jR_j - LPp_j$

where $R_j = LP\sum_{k=j+1}^{N} p_k$

where $S_j = (1-t)[i_jC_{j-1} - r_jp_jLP] - tT_j$ Letting $D_j = \sum_{k=1}^{j} S_k$ we get

$D_j - D_{j-1} = (1-t)[i_jC_{j-1} - r_jp_jLP] - tT_j$

$= (1-t)\left[i_j\left(C_0 + D_{j-1} - LP\sum_{k=1}^{j-1} p_k\right) - r_jp_jLP\right] + tLPp_j - t\Delta R_j$ (where $\Delta R_j = d_{j-1}R_{j-1} - d_jR_j$)

$= (u_j - 1)D_{j-1} + \lambda_j$ if we let λ_j represent all of the terms that do not involve D and $u_j = 1 + (1-t)i_j$

Hence if we multiply both sides of this equation by $\prod_{m=1}^{j} u_m^{-1}$ and sum we get

$\sum_{j=1}^{n}\left(D_j\prod_{m=1}^{j} u_m^{-1} - D_{j-1}\prod_{m=1}^{j} u_m^{-1}u_j\right) = D_n\prod_{m=1}^{n} u_m^{-1} = \sum_{j=1}^{n}\lambda_j\prod_{m=1}^{j} u_m^{-1}$ because $D_0 = 0$ by its definition. Thus

$D_n = \sum_{j=1}^{n} S_j = \sum_{j=1}^{n}\lambda_j\prod_{m=j+1}^{n} u_m = S_1^{*} - S_2^{*} + S_3^{*} + S_4^{*}$

$S_1^{*} = C_0\sum_{j=1}^{n}(u_j - 1)\prod_{m=j+1}^{n} u_m = C_0\sum_{j=1}^{n}\left(\prod_{m=j}^{n} u_m - \prod_{m=j+1}^{n} u_m\right) = C_0\left(\prod_{m=1}^{n} u_m - 1\right)$

$S_2^{*} = LP\sum_{j=1}^{n}(u_j - 1)\prod_{m=j+1}^{n} u_m\sum_{k=1}^{j-1} p_k = LP\left[\sum_{k=1}^{n-1} p_k\left(\prod_{m=k+1}^{n} u_m - 1\right)\right] = LP\left(\prod_{m=1}^{n} u_m\right)A_{n-1}$

$- LP\sum_{k=1}^{n-1} p_k$ {See Appendix 6} for $A_n = \sum_{k=1}^{n} p_k\prod_{m=1}^{k} u_m^{-1}$

$S_3^{*} = LPt\sum_{j=1}^{n}\prod_{m=j+1}^{n} u_m p_j - LP(1-t)\sum_{j=1}^{n}\prod_{m=j+1}^{n} u_m p_j r_j$

$= LPt\left(\prod_{m=1}^{n} u_m\right)A_n - LP(1-t)\left(\prod_{m=1}^{n} u_m\right)A_n^{*}$ $\left(\text{for } A_n^{*} = \sum_{j=1}^{n} r_jp_j\prod_{m=1}^{j} u_m^{-1}\right)$

$S_4^{*} = -t\sum_{j=1}^{n}\Delta R_j\left(\prod_{m=1}^{k} u_m^{-1}\right) = (1-t)LP\prod_{m=1}^{n} u_m(A_n + A_n^{*}) + \alpha_nR_nd_n - LPd_0\prod_{m=1}^{n} u_m + LP(1+r)\beta_n$

$- LP(1-t)\gamma_n(1+r_n)p_n$ {See Appendix 7}

Adding together these simplified sums we get

$$\sum_{j=1}^{n} S_j = C_0\left(\prod_{m=1}^{n} u_m - 1\right) - \left[LP\prod_{m=1}^{n} u_m A_{n-1} - LP\sum_{k=1}^{n-1} p_k\right] + \left\{LPt\prod_{m=1}^{n} u_m A_n - (1-t)\prod_{m=1}^{n} u_m A_n^{\bullet}\right\}$$

$$+ (1-t)LP\prod_{m=1}^{n} u_m\left[(A_n + A_n^{\bullet}) - d_0\right] + \alpha_n R_n d_n - tLPd_0\prod_{m=1}^{n} u_m + LP(1+r)\beta_n$$

$$-(1-t)LP\gamma_n(1+r_n)p_n$$

$$= C_0\left(\prod_{m=1}^{n} u_m - 1\right) + LP\left[t + (1-t)\right]\prod_{m=1}^{n} u_m A_n - LP\prod_{m=1}^{n} u_m A_{n-1}$$

$$+ LP\sum_{k=1}^{n-1} p_k - d_0 LP\prod_{m=1}^{n} u_m + \alpha_n R_n d_n + LP(1+r)\beta_n - (1-t)LP\gamma_n(1+r_n)p_n$$

$$= C_0\left(\prod_{m=1}^{n} u_m - 1\right) + LP\prod_{m=1}^{n} u_m\left(A_n - A_{n-1}\right) + LP\sum_{k=1}^{n-1} p_k - d_0 LP\prod_{m=1}^{n} u_m + \alpha_n R_n d_n$$

$$+ LP(1+r)\beta_n - (1-t)LP\gamma_n(1+r_n)p_n$$

$$= C_0\left(\prod_{m=1}^{n} u_m - 1\right) + LP\prod_{m=1}^{n} u_m\left(p_n\prod_{m=1}^{n} u_m^{-1}\right) + LP\sum_{k=1}^{n-1} p_k - d_0 LP\prod_{m=1}^{n} u_m + \alpha_n R_n d_n$$

$$+ LP(1+r)\beta_n - (1-t)LP\gamma_n(1+r_n)p_n$$

$$= C_0\left(\prod_{m=1}^{n} u_m - 1\right) + LP\sum_{k=1}^{n} p_k - d_0 LP\prod_{m=1}^{n} u_m + \alpha_n R_n d_n + LP(1+r)\beta_n - (1-t)LP\gamma_n(1+r_n)p_n$$

Adding this to the expression for C_n we get

$$C_n = \left(C_0 - d_0 LP\right)\prod_{m=1}^{n} u_m + \alpha_n R_n d_n + LP(1+r)\beta_n - (1-t)LP\gamma_n(1+r_n)p_n$$

where $\alpha_n = ivt + v\alpha_{n-1}u_n$ and $\beta_n = \beta_{n-1}u_n + (\alpha_n - t)p_n$ ($\alpha_0 = 1$ and $\beta_0 = 0$)

$$\gamma_n = \frac{\gamma_{n-1}(1+r_{n-1})u_n p_{n-1}}{(1+r_n)p_n} + 1 \quad (\gamma_0 = 0)$$

APPENDIX 6

Simplification of the S_2^* term in the expression for the sum of the surplus contributions

$$\text{Let } q_j = \prod_{m=j+1}^{N} u_m$$

$$LP\sum_{j=1}^{n}(u_j - 1)\prod_{m=j+1}^{n} u_m \sum_{k=1}^{j-1} p_k$$

$$= LP\sum_{j=2}^{n}\left(\frac{q_{j-1}}{q_j} - 1\right)\frac{q_j}{q_n}\sum_{k=1}^{j-1} p_k$$

$$= LP\sum_{j=1}^{n-1}\frac{(q_j - q_{j+1})}{q_n}\sum_{k=1}^{j} p_k = LP\sum_{k=1}^{n-1} p_k \sum_{j=k}^{n-1}\frac{(q_j - q_{j+1})}{q_n}$$

$$= LP\sum_{k=1}^{n-1}\left(\frac{q_k}{q_n} - 1\right)p_k$$

$$= LP\left(\prod_{m=1}^{n} u_m\right)\sum_{k=1}^{n-1} p_k\left(\prod_{m=1}^{k} u_m^{-1}\right) - LP\sum_{k=1}^{n-1} p_k$$

$$= LP\left(\prod_{m=1}^{n} u_m\right)A_{n-1} - LP\sum_{k=1}^{n-1} p_k \quad \text{where } A_n = \sum_{k=1}^{n} p_k\left(\prod_{m=1}^{k} u_m^{-1}\right)$$

APPENDIX 7

(See the defintions on the second page)

Assume that there is a closed form for $t\sum_{j=1}^{n}\frac{q_0 w_j}{q_n}\left(R_j d_j - R_{j-1}d_{j-1}\right)$ that is equal to

$$(1 - t)LP\frac{q_0}{q_n}(A_n + A_n^{\bullet}) + \alpha_n R_n d_n - LPd_0\frac{q_0}{q_n} + LP(1+r)\beta_n - (1-t)LP(1+r_n)p_n\gamma_n$$

$$t\sum_{j=1}^{n+1}\frac{q_0 w_j}{q_{n+1}}\left(R_j d_j - R_{j-1}d_{j-1}\right)$$

$$= \frac{q_n}{q_{n+1}}\left((1 - t)LP\frac{q_0}{q_n}(A_n + A_n^{\bullet}) + \alpha_n R_n d_n + LP(1+r)\beta_n - LPd_0\frac{q_0}{q_n} - (1-t)LP(1+r_n)p_n\gamma_n\right)$$

$$+ t\frac{q_0 w_{n+1}}{q_{n+1}}(R_{n+1}d_{n+1} - R_n d_n)$$

$$= (1 - t)LP\frac{q_0}{q_{n+1}}(A_n + A_n^{\bullet}) + \alpha_n u_{n+1} R_n d_n - LPd_0\frac{q_0}{q_{n+1}} + t(R_{n+1}d_{n+1} - R_n d_n)$$

$$+ LP(1+r_n)\beta_n u_{n+1} - u_{n+1}(1-t)LP(1+r_n)p_n\gamma_n$$

$$= (1 - t)LP\frac{q_0}{q_{n+1}}\left[(A_{n+1} - w_{n+1}p_{n+1}) + (A_{n+1}^{\bullet} - w_{n+1}p_{n+1}r_{n+1})\right]$$

$$+ R_{n+1}d_{n+1}\left[t - tv + v\alpha_n u_{n+1}\right] + LP(1+r)p_{n+1}(v\alpha_n u_{n+1} - vt) - LPd_0\frac{q_0}{q_{n+1}} + LP(1+r)\beta_n u_{n+1}$$

$$- LP(1-t)\left[(1+r_n)u_{n+1}p_n\gamma_n + (1+r_{n+1})p_{n+1}\right]$$

$$= (1 - t)LP\frac{q_0}{q_{n+1}}\left(A_{n+1} + A_{n+1}^{\bullet}\right) + LP(1+r)\left[p_{n+1}(\alpha_n v u_{n+1} - vt) + \beta_n u_{n+1}\right] +$$

$$R_{n+1}d_{n+1}\left[ivt + v\alpha_n u_{n+1}\right] - LPd_0\frac{q_0}{q_{n+1}} - LP(1-t)\left[(1+r_n)u_{n+1}p_n\gamma_n + (1+r_{n+1})p_{n+1}\right]$$

$$= (1 - t)LP\frac{q_0}{q_{n+1}}(A_{n+1} + A_{n+1}^{\bullet}) + LP(1+r)\beta_{n+1} + \alpha_{n+1}R_{n+1}d_{n+1} - LPd_0\frac{q_0}{q_{n+1}} - LP(1+r_{n+1})p_{n+1}(1-t)\gamma_{n+1}$$

$$\Rightarrow \alpha_{n+1} = ivt + v\alpha_n u_{n+1} \text{ and } \beta_{n+1} = p_{n+1}(\alpha_n v u_{n+1} - vt) + \beta_n u_{n+1}$$

$$\gamma_{n+1} = \frac{(1+r_n)u_{n+1}p_n\gamma_n}{(1+r_{n+1})p_{n+1}} + 1$$

If we let $\alpha_0 = 1$, $\beta_0 = 0$ and $\gamma_0 = 0$ then it is easy to verify that this equation holds for n = 1 (See below). The algebra above shows by induction that it is true for all n.

APPENDIX 7 (Continued)

The proof of the formula above for $n = 1$:

$$t\frac{q_0 w_1}{q_1}(R_1 d_1 - R_0 d_0) = t(R_1 d_1 - R_0 d_0)$$

$$= tLP\left(\sum_{j=2}^{N} v^{j-3/2} p_j - \sum_{j=1}^{N} v^{j-1/2} p_j\right)$$

$$= tLP\left[v^{-1}\left(\sum_{j=1}^{N} v^{j-1/2} p_j - v^{1/2} p_1\right) - d_0\right]$$

$$= tLP(1+i)d_0 - tLP(1+r)p_1 - tLPd_0$$

$$= itLPd_0 - tLP(1+r)p_1$$

The formula expression at $n = 1$ is:

$$(1-t)LP\frac{q_0}{q_1}(A_1 + A_1^*) + \alpha_1 R_1 d_1 + LP(1+r)\beta_1 - LPd_0\frac{q_0}{q_1} - LP(1-t)\gamma_1(1+r_1)p_1$$

$$= LP(1-t)(1+r_1)p_1 + (ivt + vu_1)[LP(1+i)d_0 - LP(1+r)p_1] - LP(1-t)\gamma_1(1+r_1)p_1 +$$
$$LP(1+r)[vu_1 - vt]p_1 - LPd_0 u_1$$

$$= itLPd_0 + LP(1+r)[vu_1 - vt - ivt - vu_1]p_1 \quad (\text{since } \gamma_1 = 1)$$

$$= itLPd_0 - tLP(1+r)p_1$$

Definitions used in the derivation above

$$R_{n+1}d_{n+1} = LP\sum_{k=n+2}^{N} v^{k-n-3/2} p_k$$
$$= LP\left[\sum_{k=n+1}^{N} v^{k-n-3/2} p_k - p_{n+1} v^{-1/2}\right]$$
$$= LPv^{-1}\left[\sum_{k=n+1}^{N} v^{k-n-1/2} p_k\right] - LPp_{n+1} v^{-1/2}$$
$$= v^{-1} R_n d_n - LP(1+r)p_{n+1} \Rightarrow$$
$$R_n d_n = vR_{n+1}d_{n+1} + LPv(1+r)p_{n+1}$$

Let $q_j = \prod_{m=j+1}^{N} u_m$ and let $w_j = \prod_{m=1}^{j} u_m^{-1}$

$$\Rightarrow \frac{q_{j-1}}{q_j} = u_j = \frac{w_{j-1}}{w_j} \Rightarrow \frac{q_0}{q_j} = \frac{1}{w_j}$$

$$\Rightarrow A_n = \sum_{j=1}^{n} w_j p_j \left(A_n^* = \sum_{j=1}^{n} r_j p_j \prod_{m=1}^{j} u_m^{-1} = \sum_{j=1}^{n} w_j r_j p_j \right)$$

APPENDIX 8

Solution to the alpha recurrence:

$$\alpha_j = \alpha_{j-1} v u_j + ivt \Rightarrow$$

$$\alpha_j v^{-j} \prod_{m=1}^{j} u_m^{-1} - \alpha_{j-1} v^{-(j-1)} \prod_{m=1}^{j-1} u_m^{-1} = ivtv^{-j} \prod_{m=1}^{j} u_m^{-1} \Rightarrow$$

$$\alpha_n v^{-n} \prod_{m=1}^{n} u_m^{-1} - 1 = it \sum_{j=1}^{n} v^{-j+1} \prod_{m=1}^{j} u_m^{-1} \Rightarrow$$

$$\alpha_n = v^n \prod_{m=1}^{n} u_m + it \sum_{j=1}^{n} v^{n-j+1} \prod_{m=j+1}^{n} u_m$$

Solution to the beta recurrence:

$$\beta_j = \beta_{j-1} u_j + (\alpha_{j-1} v u_j - vt) p_j \Rightarrow$$

$$\beta_j \prod_{m=1}^{j} u_m^{-1} - \beta_{j-1} \prod_{m=1}^{j-1} u_m^{-1} = \prod_{m=1}^{j} u_m^{-1} (\alpha_j - t) p_j \Rightarrow$$

$$\beta_n \prod_{m=1}^{n} u_m^{-1} = \sum_{j=1}^{n} (\alpha_j - t) p_j \prod_{m=1}^{j} u_m^{-1} \Rightarrow$$

$$\beta_n = \prod_{m=1}^{n} u_m \left[\sum_{j=1}^{n} \left(v^j \prod_{m=1}^{j} u_m + it \sum_{k=1}^{j} v^{j-k+1} \prod_{m=k+1}^{j} u_m \right) p_j \prod_{m=1}^{j} u_m^{-1} \right] - t \prod_{m=1}^{n} u_m \sum_{j=1}^{n} p_j \prod_{m=1}^{j} u_m^{-1}$$

$$= \prod_{m=1}^{n} u_m \left[\sum_{j=1}^{n} \left(v^j p_j + it \sum_{k=1}^{j} v^{j-k+1} p_j \prod_{m=1}^{k} u_m^{-1} \right) \right] - t \sum_{j=1}^{n} p_j \prod_{m=j+1}^{n} u_m$$

$$= \prod_{m=1}^{n} u_m \sum_{j=1}^{n} v^j p_j + it \sum_{j=1}^{n} v^j p_j \sum_{k=1}^{j} v^{-k+1} \prod_{m=k+1}^{n} u_m - t \sum_{j=1}^{n} p_j \prod_{m=j+1}^{n} u_m$$

Solution to the gamma recurrence:

Let $\varphi_k = (1+r_k) p_k$ Then

$$\gamma_k \varphi_k = \gamma_{k-1} u_k \varphi_{k-1} + \varphi_k \Rightarrow$$

$$\gamma_k \varphi_k \prod_{m=1}^{k} u_m^{-1} - \gamma_{k-1} \varphi_{k-1} \prod_{m=1}^{k-1} u_m^{-1} = \prod_{m=1}^{k} u_m^{-1} \varphi_k \Rightarrow$$

$$\gamma_n \varphi_n \prod_{m=1}^{n} u_m^{-1} = \sum_{k=1}^{n} \prod_{m=1}^{k} u_m^{-1} \varphi_k \text{ or}$$

$$\gamma_n = \frac{\sum_{k=1}^{n} (1+r_k) p_k \prod_{m=k+1}^{n} u_m}{(1+r_n) p_n}$$

Final Report of the CAS Research Project on Full Information Equity Betas for Property-Liability Insurance Including By-Line Estimates

The Risk Premium Project

August, 2003

Final Report of CAS Research Project on

Full Information Equity Betas for Property-Liability Insurance Including By-Line Estimates

Submitted by

The Risk Premium Project (RPP)

Academic Investigators	**Industry Investigators**
J. David Cummins, Ph.D.	Richard A. Derrig, Ph.D.
University of Pennsylvania	Automobile Insurers Bureau of Massachusetts
Richard D. Phillips, Ph.D.	Robert P. Butsic, MBA, ASA, MAAA
Georgia State University	Fireman's Fund Insurance Company

Introduction and Statement of Purpose

The Risk Premium Project was organized in the Summer of 1999 to address the question of the equilibrium valuation of property and liability insurance risks as posed by the CAS's Committee on the Theory of Risk (COTOR). Phases I and II of the project were completed in April 2000 with a follow-up empirical Phase III. The goal of Phase III is to demonstrate the appropriateness of the pricing methodologies we discussed theoretically in Phases 1 and II and to develop procedures and parameterized models necessary to implement the methodologies empirically. A significant outcome of the research we propose will be the practical application of the results for use in the pricing and capital budgeting models currently being used by insurers.

This report is organized as follows: Section 1 provides a brief review of Phases I and II and highlights the theoretical conclusions. Further details can be found in the interim report published in the *Forum of the Casualty Actuarial Society* (Fall 2000) or on the CAS website at www.casact.org/cotor/rpp.htm. Sections 2 describes the empirical investigation on equity betas, including a statement regarding the importance and relevancy of the research, the data sources and the empirical methodologies we utilize.

1. The Risk Premium Project: A Review of Phases I and II

The goal of Phase I of the Risk Premium Project was to identify and review the actuarial and financial literature relating to the question of the equilibrium valuation of property-liability risks with a primary focus on research that has been published in the last 10-15 years. In all, over two hundred research papers and books were considered. The final bibliography contains 138 references from 37 different sources that in our view were deemed especially relevant to the valuation of insurance liabilities. All references were assigned to one of several thematic categories and the entire set of RPP references can be search online at www.casact.org/cotor. The references appear as an annotated bibliography searchable by author, subject, or keyword. Full PDF versions of many of the papers in the bibliography, with and without annotations, are also available on the web site via publisher links.

In Phase II of the project we attempted to synthesize the literature to provide a compact discussion of the theoretical conclusions that can be reached based upon the most recent evidence. Although we do not provide the details here, the five principal conclusions that have a direct bearing on the research question are shown below:

I. The opinions of financial economists and actuaries regarding the role of systematic vs. non-systematic risks in determining equilibrium insurance prices are converging. Both see a role for non-systematic risk in pricing although an estimate of the non-systematic cost of risk is not yet well known.

II. It is well known that a the systematic risk adjustment for the cash flows associated with a line of insurance should be included in the discount rate used to determine the fair value of the insurance premium.

However, it is not well know that in addition to the adjustment for common innovations between cash flows and market returns, the discount rate will also be a function of the maturity structure of the cash flows. This result suggests the equilibrium valuation of a line of insurance will require a charge for systematic risk even though the liability cash flows are triggered by events largely uncorrelated with market returns or other macroeconomic factors.

III. The returns of financial assets cannot be adequately explained by the CAPM beta. Researchers have shown extensions of the CAPM which include additional factors significantly enhance the explanatory power of the models. In addition, although research using more sophisticated empirical tests has been published extending the CAPM, similar research focusing on insurance company returns does not currently exist.

IV. A theoretically consistent way to allocate the costs of holding equity capital to individual lines of insurance has been identified. Thus, the costs associated with holding capital can now be charged to individual lines of insurance.

V. The risk of insurer default to the policyholder should be recognized in pricing the risk transfer.

Research into some of the conclusions we list above is still ongoing and in some cases the reasons why some of the relationships exist have not been fully explained. In addition, empirical tests of many of the theoretical conclusions are either currently underway or have not yet been conducted. That being said, it is fair to say that a coherent approach to the valuation of insurance liabilities is emerging and much progress has been made that should be exploited to investigate the issues of equilibrium insurance pricing and insurance liability discounting. The empirical investigation we discuss below takes advantage of the recent research and should prove useful for both industry professionals seeking to more efficiently price their products as well as academics seeking to understand and evaluate the functioning of this important component of the financial services industry.

Phase III, Project 1: Full Information Equity Betas for Property-Liability Insurance Including By Line Estimates

Importance of the Research

Modern financial theory suggests the identification of systematic risk is an important prerequisite to determine the equilibrium prices of securities issued by corporations, the valuation of privately held companies, and, more generally, in the pricing of goods and services offered by firms in regulated industries. The application of this theory to insurance pricing (e.g., Myers and Cohn 1987) has required an estimate of the insurer's equity CAPM beta which can then be used to price the insurer's liabilities as a residual.

Unfortunately prior research investigating the equity cost of capital for insurers, and therefore estimates of the systematic risk of the insurer's liabilities, suffers from two distinct problems. First, recent research documents a number of empirical anomalies that can not be reconciled within the assumptions that underlie the Capital Asset Pricing Model. Most notably, Fama and French (1992) show investments in small-cap stocks appear to earn average returns *higher* than would otherwise be predicted by the CAPM even after controlling for beta. In addition, assets with high book-to-market equity ratios (value stocks) have *higher* average returns after accounting for market beta. Although other factors have been studied (including leverage, dividend yield, earnings/price ratio, etc.) the dominant multifactor model to date is the Fama-French three-factor model. Although the failures of the CAPM are well documented in the literature, the implications for the cost of capital estimates in the insurance have not been identified. Thus, the first goal of this project is to provide estimates of the cost of capital for the insurance industry recognizing the weaknesses of the traditional application of the CAPM.

The second primary limitation of the application of modern capital budgeting model to price insurance risks is the general lack of reliable information regarding differences in the systematic risk across the various lines of insurance. Prior research seeking to estimate liability betas by line of insurance has met with little success.

Cummins and Harrington (1985) adopted an accounting based approach and estimated underwriting betas using based upon accounting data. They report the estimates were highly unstable – a result entirely consistent with other papers in this literature that report accounting betas are typically not highly correlated with market betas. Cox and Griepentrog (1988) adapted the pure-play approach of Fuller and Kerr (1981) to estimate divisional costs of capital for insurers but report the resulting cost of capital estimates were also unreliable. In addition, the estimates the authors do report are likely biased upwards since the pure-play approach requires the elimination for firms conducting business in more than one line of insurance. This requirement leaves insurers that write in only one line insurance who tend to be relatively low-market-capitalization firms whose costs of capital are known to be higher due to the small cap stock effect documented by Fama and French.

Completed Research

In this project we overcame the limitations of both the pure play and accounting based approaches and estimate full information equity betas, with autocorrelation adjustments (sum betas), for a typical property-liability insurer in a manner similar to Kaplan and Peterson (1997). The full information approach assumes the systematic risk of a conglomerate firm is a weighted average divisions of the firm and has gained widespread attention as it is a market based approach that allows the one to estimate industry costs of capital without eliminating so many firms from the analysis.

The estimation technique extends the Kaplan and Petersen approach in two important ways. First, we estimate the firm specific betas using the three-factor model proposed by Fama and French (1992, 1997). Thus, our estimates will more closely reflect the state-of-the-art in cost of capital estimation techniques. Second, we propose to extend their analysis beyond industry estimates of the cost of capital to estimate the equity cost of capital by line or groups of lines of insurance. Thus, the results on relative risk will be useful for actuaries pricing various lines of insurance and should also shed light on the role the duration of the liability cash flows has in determining the systematic risk of the line of insurance. Estimation for individual lines of insurance proved to be unstable due primarily to data limitations. Consequently, final estimates are made for portfolios of lines of insurance: short-tail and long-tail, personal and commercial, and auto, workers' compensation and all other lines.

The principal document constituting our report is the enclosed academic style paper by Cummins and Phillips: Estimating the Cost of Equity Capital for Property-Liability Insurers, dated June 23, 2003. This paper has been submitted to the Journal of Risk and Insurance for peer review and, when published will cite the CAS as the financial sponsor. All members of the RPP acknowledge the generous commitment of time by the members of the Committee on the Theory of Risk. We look forward to the completion of the companion IRC Phase III project on allocation of capital.

Estimating the Cost of Equity Capital For Property-Liability Insurers

J. David Cummins and Richard D. Phillips

ESTIMATING THE COST OF EQUITY CAPITAL FOR PROPERTY-LIABILITY INSURERS

June 23, 2003

By

J. David Cummins
Insurance and Risk Management Department
University of Pennsylvania
3641 Locust Walk
Philadelphia, PA 19104
tel: 215-898-5644
fax:215-898-3589
cummins@wharton.upenn.edu

Richard D. Phillips*
Department of Risk Management and Insurance
Georgia State University
P.O. Box 4036
Atlanta, GA 30302
tel: 404-651-3397
fax: 404-651-4219
rphillips@gsu.edu

* Corresponding author. Preliminary - please do not cite without first checking with the authors for a later version.

Estimating the Cost of Equity Capital for Property-Liability Insurers

Abstract

This paper presents new evidence on estimates of the cost of equity capital by line of insurance for the property-liability insurance industry. To do so we obtain firm beta estimates and then use the recently developed full-information industry beta methodology to decompose the cost of capital by line. We obtain beta estimates using both the standard one-factor CAPM model as well as the Fama-French three-factor cost of capital model. The analysis suggests the cost of capital for insurers using the Fama-French model are significantly higher than estimates based upon the CAPM. In addition, we find evidence of significant differences in the cost of equity capital across lines, indicating that the use of a single company-wide cost of capital is generally not appropriate.

1. Introduction

Cost of capital estimation is becoming increasingly important for property-liability insurers. First introduced during the 1970s in regulatory proceedings in New Jersey, Massachusetts, and other states, the application of financial methods in pricing, reserving, and other types of financial decision making has grown rapidly over the past two decades.[1] Recent developments include the emergence of asset-liability management (ALM) techniques (Panjer 1998), the development of methodologies to allocate equity capital by line of business (e.g., Myers and Read 2001), the increased focus on market-based project evaluation techniques such as risk-adjusted return on capital (RAROC), and the projected introduction of fair value accounting for insurer liabilities (Girard 2002). These and other changes have intensified the need to find reliable methods to estimate the cost of capital for insurance firms.

In terms of financial theory, insurers are no different from other corporations in the economy with respect to the general factors that determine the cost of capital and the market value of the firm. However, it is well-recognized that the cost of capital varies across industries due to the heterogeneity of the risks facing firms in various sectors of the economy; and cost of capital research has shown that there is a significant industry factor for insurance (Fama and French 1997). Insurance is a very diverse

[1]Most of the important early financial pricing papers are collected in Cummins and Harrington (1987). More recent developments are reviewed in Cummins and Phillips (2001). Insurance pricing models that rely on estimates of the cost of capital are presented in Cummins (1990) and Taylor (1994).

industry, however, encompassing numerous lines of business with different risk characteristics. It is unlikely, for example, that the cost of capital for a firm specializing in life insurance will be the same as the cost of capital for a firm emphasizing workers' compensation or commercial liability insurance. Unfortunately, little progress has been made in estimating costs of capital for insurers with different business line compositions. The objective of the present paper is to remedy this deficiency in the existing literature by developing cost of capital models that reflect the line of business characteristics of firms in the property-liability insurance industry. We focus on property-liability insurers because costs of capital tend to differ across industry segments (see below) and because of the long-term interest in cost of capital for property-liability insurers due to price regulation, which is not present for other types of insurance.

The problem addressed in this paper, i.e., the estimation of the cost of equity capital for insurers with different business line compositions, has been studied in the financial literature as the problem of estimating the cost of capital for divisions of conglomerate firms. The problem in estimating the divisional cost of capital is that the conglomerate firm itself rather than the division is traded in the capital market. Thus, it is possible to use market value data to estimate the overall cost of capital for the conglomerate but not for the individual divisions comprising the firm. The classic approach for estimating the divisional cost of capital is the *pure play approach* (Fuller and Kerr 1981). The pure play technique involves identifying publicly traded firms that specialize in the same product as the division under consideration (i.e., that have a "pure play" in that product) and then approximating the divisional cost of capital as the average cost of capital for the pure play firms.

Although the pure play technique performs well in some instances, particularly when a relatively large number of pure play firms of various sizes can be found, there are also situations where it does not provide a satisfactory solution to the divisional cost of capital problem. In particular, there may be only a few true specialist firms in some product lines and they may be small, accounting for only a small proportion of industry output (Ibbotson Associates 2001). Because small firms tend to have higher costs of capital than large firms, using pure play cost of capital estimates from relatively small firms to determine the cost of capital of a (potentially much larger) division of a conglomerate firm can lead to biased estimates of the divisional cost of capital. Unfortunately, the property-liability insurance industry is an example of the type of industry where the pure play approach is unlikely to work very well. The vast majority of insurance premiums are written by firms that write several lines of business rather than specializing in one or two closely-related lines of business. In addition, relatively few insurers are publicly traded, with the majority of firms in the industry owned by financial or non-financial conglomerates or having the mutual ownership form.

In this paper, we adopt a relatively new methodology, the *full-information industry beta (FIB) approach*, that overcomes the principal limitations of the pure-play methodology. The full-information beta methodology was first proposed by Ehrhardt and Bhagwar (1991) and significantly modified by Kaplan and Peterson (1998). Instead of discarding the overall cost of capital estimates for conglomerates, as is done in pure play analysis, the full-information beta approach utilizes a sample of conglomerate and specialist firms to identify the impact of various lines of business on the cost of capital. The underlying insight is that the *observable* beta for the conglomerate is a weighted average of the *unobservable* betas of the underlying lines of business. The method proceeds by performing a cross-sectional regression for a sample of firms, where the dependent variable is the observable beta and the independent variables measure the firms' participation in various industries and lines of business. The coefficients of the line of business participation variables are then interpreted as the full-information beta coefficients for the business lines. The resulting regression equation can be used by firms outside of the estimation sample to estimate the cost of capital taking into account their own line of business compositions. Hence, the results could be used to produce cost of capital estimates for non-traded stock firms and mutuals. The method also can be used to estimate the cost of capital for insurers specializing in particular industry segments or for subsidiaries of insurers and conglomerates specializing in various lines of business.

The approach taken in this paper is to illustrate the full-information beta approach to cost of capital estimation using a sample consisting of all firms (insurance and non-insurance) listed in the Compustat data base that meet our sample selection criteria for the sample period 1997-2000. Our first set of FIB regressions, based on the Compustat data, produces beta estimates for all two-digit industries defined by the North American Industry Classification System (NAICS), including the property-liability insurance industry. Our second set of FIB regressions supplements the Compustat data with data from the National Association of Insurance Commissioners (NAIC) in order to estimate the cost of capital by line of property-liability insurance. To illustrate the application of the FIB technique in estimating line of business betas, we estimate models that provide cost of capital estimates for heavily regulated lines of business (automobile versus workers compensation insurance), short-tail versus long-tail lines of business, and personal versus commercial lines. For comparison with the FIB estimates, we also estimate the cost of capital for property-liability insurers using the pure play approach.

The betas used as the dependent variables in our full-information beta regressions come from two asset-pricing models – the capital asset pricing model (CAPM) and the three-factor model developed by Fama and French (1992, 1993, 1997). The CAPM and Fama-French methods were chosen for analysis because they are used frequently in determining the cost of capital for corporate capital budgeting and regulatory purposes. The CAPM is important because it was the first equilibrium asset pricing model developed by financial theorists and still plays a prominent role in a many practical applications (Graham and Harvey 2001). The Fama-French three-factor model (hereinafter the FF3F model) was developed in response to the criticism that the CAPM tends to give inaccurate estimates of the cost of capital because it omits important financial risk-factors. The FF3F model retains the CAPM's single factor for systematic market risk and adds factors to capture the effects of firm size (defined in terms of total market capitalization) and the ratio of the book value of equity (BV) to the market value of equity (MV). The former factor controls for the "small firm effect," i.e., the observation that the cost of capital is inversely related to firm size. The BV to MV ratio reflects financial distress, with financially vulnerable firms having higher values of this ratio than stronger firms. This factor controls for the tendency of investors to require higher expected returns on stocks in financially vulnerable firms since these firms will perform particularly poorly exactly when individual investors' portfolios are experiencing overall losses.[2] In implementing the CAPM, there is one FIB regression, with the CAPM beta coefficient as the dependent variable. In the Fama-French methodology, there are three FIB regressions, one each for the three risk-factor coefficients in the Fama-French model.

There have been several prior papers on cost of capital estimation for property-liability insurers. Cummins and Harrington (1985) utilize quarterly profit data to estimate the cost of capital for a sample of fourteen property-liability insurers covering the late 1970s and early 1980s. Their beta estimates were somewhat unstable and conformed to the CAPM in the 1980s but not in the 1970s. Cox and Griepentrog (1988) implemented the Fuller-Kerr (1981) pure play technique for a sample of 26 to 31 insurers (depending on the year) using data from the mid-1970s. Like Cummins and Harrington's, their estimates of the cost of capital tend to be somewhat unstable. Cummins and Lamm-Tennant (1994) develop theoretical and empirical models showing that insurer costs of capital are related to leverage (the ratios of policy reserves-to-assets and the ratio of financial debt-to-assets). Their results suggest that long-tail commercial lines of property-liability insurance tend to have higher costs of capital than short-tail lines. Their sample period ends in 1989. Lee and Cummins (1998) estimate the cost of equity capital for property-liability insurers using the CAPM, the arbitrage pricing theory (APT) model, and a unified CAPM-APT model developed by Wei (1988). They find that the APT and the Wei model perform better than the CAPM in forecasting the cost of capital for insurers. Except for Lee-Cummins, none of the prior research uses data after the 1980s, and none except Cummins and Lamm-Tennant attempts to estimate the cost of capital by line.

The present paper improves on the existing literature by being the first to investigate both the

---

[2]The construction of the size and book to market factors are defined in more detail below. Cochrane (1999) reviews the recent empirical asset pricing literature and provides an intuitive discussion of the non-diversifiable risks proxied by the size and financial distress risk factors.

Fama-French three-factor model and the full-information beta technique to study the insurance industry. We also innovate by conducting the first application in any industry of the full-information beta technique to explain the factors of the Fama-French three-factor model. Our research improves on the prior insurance cost of capital literature by using more recent data and focusing on the relationship between the cost of capital and line of business composition. In addition, the sample size in the present study is larger than the samples used in most of the prior research on property-liability insurer costs of capital.

The remainder of the paper is organized as follows: Methods for estimating the cost of capital are discussed in section 2. Section 3 discusses our sample and methodology. The results are presented in section 4, and section 5 concludes.

2. Cost of Capital Methodologies

Several important methodologies have been developed to estimate the cost of capital for corporate financial decision making and regulatory purposes. In this section, we outline these methodologies and explain our rationale for focusing on the approaches considered in the empirical section of the paper. Because the models are well-known, the discussion is brief, and the reader is referred to the literature for further discussion.[3] The details on the implementation of the models are presented in section 3.

The Discounted Cash Flow (DCF) Model

The discounted cash flow (DCF) model has been widely used in regulatory proceedings, both in the insurance and the public utility industries.[4] The model is based on the financial-theoretic proposition that the value of any asset is the present value of its cash flows, where the discount rate is

[3]An overview of the models is presented in Ibbotson Associates (2001) and the technical details are provided in Campbell, Lo, and MacKinlay (1997).

[4]Brealey and Myers (2002) and Ibbotson Associates (2001) provide details on several variants of the model. Kolbe, Read and Hall (1984) and Thompson (1991) discuss the use of the models in rate regulatory proceedings. Derrig (1990) discusses the development and use of financial pricing models in the context of insurance rate regulation.

the appropriate cost of capital for the firm or project under consideration. Solving the discounted cash flow formula for the cost of capital yields the following equation:

$$r_i = \frac{C_{i1}}{V_i} + g_i \qquad (1)$$

where r_i = cost of capital for firm i,

C_{i1} = the firm's expected cash flow next period,

V_i = the market value of the firm in the current period, and

g_i = the expected growth rate of the firm's cash flow in the future.

The cash flow to market value ratio (C_{i1}/V_i) is usually approximated by the firm's projected dividend-to-price ratio or earnings-to-price ratio obtained from financial reporting services such as Thomson Financial or Value-Line. The projected growth rate, also in earnings or dividends, usually is based on a average of financial analysts' forecasts obtained from a source such as the Institutional Brokers Estimation Service (IBES). Because it is often viewed as unrealistic to project a single growth rate in perpetuity, variants of the DCF model have been developed using two or three growth rates rather than the single growth rate shown in equation (1) (Brealey and Myers 2002).

The DCF model provides a practical way to obtain cost of capital estimates that is often useful in checking the robustness of results obtained from single or multiple factor asset pricing models. However, its usefulness is limited in the insurance context because of its reliance on analysts' earnings forecasts to obtain growth rates. Analysts' earnings are typically only available for the largest publicly traded firms. Thus, the sample of companies that could be analyzed using this method is much more limited than the sample used in this study. In addition, the method does not readily lend itself to the decomposition of the cost of capital by line.

The Capital Asset Pricing Model (CAPM)

The capital asset pricing model (CAPM) has been widely used in many financial applications. The CAPM cost of capital is given by the following formula:

$$E(r_i) = r_f + \beta_{mi}[E(r_m) - r_f] \qquad (2)$$

where $E(r_i)$ = the CAPM cost of capital for firm i,

r_f = the expected return on a default risk-free rate asset,

$E(r_m)$ = the expected return on the market portfolio, and

β_{mi} = firm i's "beta coefficient" for systematic market risk = $Cov(r_i, r_m)/Var(r_m)$

The underlying rationale for the CAPM is that expected rates of return on assets traded in frictionless and informationally efficient capital markets are sufficient to compensate investors for the time value of money at the default-risk-free rate of interest plus a risk premium to compensate investors for bearing systematic market risk. The latter component of the return is equal to the firm's beta coefficient multiplied by the expected market risk premium. In applications of the CAPM, the expected market risk premium is proxied by the long-term average of the difference between the return on a broad market index and the return on a risk-free asset. The most commonly used averaging period is 1926-present, and the broad market index is a market index such as the Standard & Poor's 500 Stock index (S&P 500), the New York Stock Exchange (NYSE) composite index, or a broad market index consisting of NYSE, America Exchange, and Nasdaq stocks.[5] A U.S. government bill or bond rate is used to represent r_f, with the choice of rate usually depending upon the time horizon of the firm or project being evaluated.[6]

Beta coefficients for individual stocks are estimated by a time series regression of a firm's stock returns on the returns on the market portfolio. Most applications utilize monthly data over a five-year period to conduct the regressions. The regressions are usually estimated using ordinary least squares (OLS) with an adjustment to allow for the phenomenon of regression towards the mean, i.e., the

---

[5]Although using the long-term average of the difference between the return on a broad market index and the risk-free rate to represent the expected market risk premium is the most widely accepted approach in the literature, other approaches have been suggested. For a review of the alternatives, see Derrig and Orr (2003). The models estimated in this paper could easily be adapted to incorporate other proxies for the market risk premium.

[6]For example, Ibbotson Associates, a widely-accepted source for cost of capital estimation, recommends the 30-day U.S. Treasury bill yield for short-horizon projects, the 5-year U.S. Treasury coupon note yield for intermediate-horizon projects, and the 20-year U.S. Treasury coupon bond yield for long-horizon projects. Ibbotson Associates (2001), p. 43.

tendency of unusually high or low beta estimates to be biased away from their true expected values.[7]

The CAPM is included in our empirical analysis because it remains the most widely used asset pricing model in a wide range of practical applications such as capital budgeting and investment portfolio analysis (see, for example, Graham and Harvey 2001). In addition, more recent methods such as the FF3F method and the full-information beta technique can be viewed as generalizations or extensions of the CAPM. Thus, it is important to include the CAPM as a benchmark methodology in this study.

The Fama-French Three-Factor (FF3F) Model

As mentioned above, the Fama-French three-factor model retains the CAPM risk-premium for systematic market risk but adds risk premia for two additional factors to capture the effects of firm size and financial distress. The factors are defined in detail in section 3 below. The FF3F model has been tested extensively and shown to be a significant improvement over the CAPM (see Fama and French 1992, 1993, 1997, Schink and Bower 1994, Wang 2003).

The FF3F formula for the cost of capital is the following:

$$E(r_i) = r_f + \beta_{mi}[E(r_m) - r_f] + \beta_{si}\,\pi_s + \beta_{vi}\,\pi_v \qquad (3)$$

where β_{si} = firm i's beta coefficient for the size factor,

π_s = the expected market risk premium for firm size,

β_{vi} = firm i's beta coefficient for the financial distress factor, and

π_v = the expected market risk premium for financial distress.

The risk-premium for systematic market risk, $E(r_m) - r_f$, in the FF3F model is usually the same estimate that is used for the CAPM. This model also uses factors representing *size excess returns* and *financial distress excess returns*, where firm size is defined in terms of total market capitalization (number of shares multiplied by share price) and financial distress is proxied by the ratio of the book value of equity

[7]For a discussion of adjusting historical betas for regression towards the mean, see Elton and Gruber (1991).

(BV) to the market value of equity (MV). The excess return series are obtained monthly, and long-term averages of the returns are used to compute the risk premia, π_s and π_v. The market, size, and financial distress excess returns are used in a regression analysis to estimate the beta coefficients for systematic market risk, firm size, and financial distress. Finally, the estimated beta coefficients for each firm and the time-averaged risk premia are inserted into equation (3) to estimate the cost of capital for the firms in the sample.[8]

The Full-Information Industry Beta (FIB) Method

As mentioned above, the objective of the full-information beta methodology is to produce cost of capital estimates that reflect the line of business composition of the firm. Such estimates can be used by non-traded stock insurers and mutuals to estimate the cost of capital and can also be used to estimate the cost of capital for divisions of the firm.

The underlying premise of the FIB methodology is that the firm can be envisioned as a portfolio of assets, where the assets represent divisions, individual lines of business, or separate projects undertaken by the firm. In this conceptualization, the firm's overall market beta coefficients are weighted averages of the beta coefficients of the separate divisions of lines of business. In theory, the weight on each divisional or line of business beta is the market value of the division divided by the market value of the firm as a whole. However, because individual business units are not publicly traded,

[8]We also considered estimating multi-factor models based upon arbitrage pricing theory (APT) as developed by Ross (1976). Like the FF3F model, APT is less restrictive than the CAPM in that it allows for multiple factors rather than a single market factor. Unlike the CAPM, it does not require capital market equilibrium, only the absence of arbitrage opportunities. The APT formula has the same form as equation (3), i.e., it is the sum of a series of beta coefficients multiplied by risk factors. There is no set number of factors and the factors may or may not include the CAPM premium for systematic market risk. Although APT is an important model, we do not estimate APT costs of capital in this paper. Among the practical limitations of the APT is that it places heavier demands on the data than the CAPM or FF3F methods and thus is more difficult to implement for industries such as insurance where the number of traded stocks is relatively low. In addition, the factors that comprise empirical versions of the arbitrage pricing model are often difficult to interpret economically (i.e., they are developed using factor analysis rather than being determined a priori by theoretical arguments), and the factor risk premia are often difficult to estimate. Moreover, the generalization of the methodology to incorporate industry factors also is not straightforward. Further information on APT can be found in Roll and Ross (1980), Ingersoll (1987), and Campbell, Lo, and MacKinlay (1997). Lee and Cummins (1998) present a recent application of APT to insurance stocks.

it is not possible to use market value weights. Instead, Kaplan and Peterson (1998) recommend using divisional or line of business sales data to represent participation in different industries or activities. We adopt the Kaplan-Peterson approach in this paper.

Specifically, we seek to decompose the overall market beta coefficient (for the CAPM) or coefficients (for the FF3F model) into separate beta coefficients for each industry/line of business in which firms participate. The decomposition is accomplished by performing a cross-sectional regression for a sample of firms with the overall market beta as the dependent variable and a series of weights proxying for the firm's participation in various lines of business as explanatory variables. For example, the regression equation for the CAPM beta is:

$$\beta_i = \sum_{j=1}^{J} \beta_{fj}\, \omega_{ij} + \nu_i \tag{4}$$

where β_i = the firm's overall market beta coefficient,

β_{fj} = the full-information beta for industry j,

ω_{ij} = firm i's industry participation weight for industry j, and

ν_i = random error term for firm i.

Analogous regressions are used to estimate full-information betas for the FF3F method. The ω_{ij}, j = 1, 2, . . ., J for firm i, which sum to 1.0, reflect the firm's relative exposure to risks from each line of business. The β_{fj}, which vary by industry but not by firm, are designed to capture the impact that any particular line of business is expected to have on the overall riskiness and hence the beta coefficient of the firm. The key idea reflected in the FIB technique is that equation (4) can be used "out of sample" to estimate the overall beta coefficients β_i for non-traded firms or for individual divisions or lines of business. For example, a firm with 100% of its revenues in industry (line of business) j would have an estimated overall beta coefficient: $\beta_i = \beta_{fj}$; and a firm with 50% of its revenues from industries j and k would have an overall beta coefficient: $\beta_i = 0.5(\beta_{fj} + \beta_{fk})$.

3. Data and Methodology

Data and Sample Selection

To estimate the CAPM, FF3F, and full-information costs of capital, we need data on stock returns and insurer revenues by line of business. This section describes the data sources, sample selection procedures, and data screens employed to construct our sample.

The stock return data for the study were obtained from the University of Chicago's Center for Research on Securities Pricing (CRSP). The data consist of returns on all NYSE, AMEX, and Nasdaq stocks. We obtained the CRSP data for the period 1992-2000, permitting us to estimate costs of capital for the period 1997-2000, because we follow the standard procedure of using 60 monthly observations to estimate the parameters of our cost of capital models. In selecting the stock returns, we employed screening rules that also are standard in the cost of capital estimation literature (e.g., Fama and French 1992, 1997), i.e, we eliminated firms with estimated CAPM beta coefficients greater than 5 in absolute value and also eliminated firms that did not have at least 36 consecutive months of return information prior to June of each year of the estimation period 1997-2000. Excess return data for market systematic risk, size, and financial distress were obtained from Kenneth French's website.[9]

To obtain revenues by line of business, we utilize Compustat's Business Information File (BIF) and data from the National Association of Insurance Commissioners (NAIC). Compustat includes revenue data for firms in various industries, where industries are categorized using the North American Industry Classification System (NAICS). Insurance is included in the finance sector, which has two-digit NAICS code of 52. Within the finance sector, the NAICS system categorizes revenues by industry and sub-industry. There are several insurance sub-categories including property/casualty insurance (NAICS code 524126), property/casualty reinsurance (NAICS code 52413 or 524130), life insurance (NAICS code 524113), and health insurance (NAICS code 524114), and we utilize Compustat revenue data for these insurance lines of business in estimating our models. Within the property/casualty industry, the NAICS system does not further categorize revenues by line of insurance. Consequently, we obtained

---

[9]The url is: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french.

data on insurance revenues by line from the NAIC annual statement CD-ROMs to supplement the Compustat data.

Estimation Methodology

Our CAPM estimates were obtained using standard procedures, described below; and our estimation approach for the FF3F model follows Fama and French (1992, 1993, 1997). Our estimation approach for the full-information beta (FIB) analysis follows Kaplan and Peterson (1998).

Cost of capital estimation using the CAPM, FF3F, and FIB analyses is conducting using a two-stage approach. In the first stage, returns on specific stocks in the sample are regressed on a market risk factor or factors to obtain separate beta coefficients for each firm. The beta estimation stage may involve more than one step, as in the case of FIB analysis. In the second stage, the beta coefficients are inserted into equations such as (2) and (3) along with the estimated market risk premia to obtain cost of capital estimates for each firm. In the following, we describe more precisely the estimation of cost of capital using the three methods and then discuss our application of the pure play approach.

The CAPM. The easiest of the methodologies to implement is the CAPM. The first stage regression is:

$$r_{it} - r_{ft} = \alpha_i + \beta_{mi}(r_{mt} - r_{ft}) + \varepsilon_{it} \qquad (5)$$

where β_{mi} = the CAPM beta coefficient for firm i ,

r_{it} = the return on stock I in period t,

r_{ft} = the risk-free rate in period t (30-day Treasury bill yield),

r_{mt} = the return on the market portfolio in period t, and

ε_{it} = a random error term for stock i.

The regression produces estimates of the parameters α_i and β_{mi}. The regression sample periods consist of 60 months of data ending in June of each year 1997 through 2000. Firms with less than 60 months of data are included in the analysis as long as they have at least 36 months of data prior to June of each year of the estimation period 1997-2000.

In practice, equation (5) works well for relatively large stocks that are frequently traded. However, it tends to give biased estimates of β_{mi} for stocks that trade infrequently, e.g., stocks that may go for days or weeks at a time with little or no trading. In order to correct for the bias created by infrequent trading, we utilize the *sum-beta approach* that has become standard in this type of analysis (e.g., Scholes and Williams 1977; Dimson 1979). Specifically, we augment the explanatory variables in equation (5) by adding the lagged value of the excess market return variable and estimate the following equation:

$$r_{it} - r_{ft} = \alpha_i + \beta_{mi0}(r_{mt} - r_{ft}) + \beta_{mi1}(r_{m,t-1} - r_{f,t-1}) + \varepsilon_{it} \qquad (6)$$

The estimated beta coefficient from this model is the sum of the contemporaneous and lagged beta estimates from equation (6), i.e., the estimated sum beta coefficient is: $\hat{\beta}_{mi} = \hat{\beta}_{mi0} + \hat{\beta}_{mi1}$.[10]

The market excess return factor (r_{mt} - r_{ft}) in equations (5) and (6) is defined as the value-weighted NYSE/AMEX/Nasdaq return less the 30 day T-bill rate lagged one month. This is the standard approach adopted in most cost of capital analyses. The estimated sum beta coefficient from equation (6) is then inserted into equation (2) to obtain the CAPM cost of capital. The market premium for systematic risk, $E(r_m)$ - r_f, is estimated as the value-weighted excess return on NYSE/AMEX/Nasdaq stocks from 1926 until June of 2000. We use the same market risk premium in all estimations to focus attention on the differences between methodologies, conditional on the risk premium, rather than varying the risk premium across the estimation period.

The Fama-French 3-Factor Model. The first-stage regression in the Fama-French (FF3F) methodology is a generalization of the CAPM regression:

$$r_{it} - r_{ft} = \alpha_i + \beta_{mi}(r_{mt} - r_{ft}) + \beta_{si}\,\pi_{st} + \beta_{vi}\,\pi_{vt} \qquad (7)$$

---

[10]Scholes and Williams (1977) and Dimson (1979) include lead terms in addition to the lagged terms in adjusting for non-synchronous trading. However, these lead terms are not necessary under the assumption that the market return is not contaminated by stale prices and are generally not included in empirical analysis similar to ours. In addition, the methodology used above is consistent with sum beta approach utilized in practice by Ibbotson Associates (see Ibbotson 2001 for further discussion).

The returns on stock i, the risk free return, and the return on the market portfolio are the same as in the CAPM regression. As mentioned above, the model also utilizes excess return series representing market excess returns for size and financial distress.

The excess returns for size and financial distress are estimated using the procedure developed in Fama and French (1992).[11] The procedure involves the formation of six portfolios utilizing all stocks traded on the New York Stock Exchange (NYSE), the American Exchange (AMEX), and the Nasdaq. The portfolios are defined as of the end of June in each year of the estimation period, 1997-2000. To estimate the excess return series for firm size, stocks are divided into two portfolios depending upon whether their market capitalization is above or below the median market capitalization for NYSE stocks. These are referred to as the high market capitalization (HMC) and low market capitalization (LMC) portfolios.[12] Likewise, to estimate the expected market risk premium for financial distress, stocks are sorted by their BV/MV ratios and divided into three portfolios depending upon whether their BV/MV ratios are in the range of the top 30 percent of NYSE stocks' BV/MV ratios, the middle 40 percent of NYSE stocks' BV/MV ratios, or the bottom 30 percent of the NYSE stocks' BV/MV ratios. We refer to these three categories as high book-to-market value (HBMV), medium book-to-market value MBMV), and low book-to-market value (LBMV) stocks.

After placing all NYSE, AMEX, and Nasdaq stocks into market capitalization and book-to-market value categories, the next step is to form six return series determined by the intersection of the two market capitalization and three book-to-market value categories, yielding six categories of stocks: (1) HMC and HBMV, (2) HMC and MBMV, (3) HMC and LBMV, (4) LMC and HBMV, (5) LMC and MBMV, and (6) LMC and LBMV. The return series are averages of the returns on the stocks in the six categories, where returns are weighted by total market capitalization of each stock to obtain weighted

---

[11]As mentioned earlier, the excess returns for systematic market risk, size, and financial distress were obtained from Kenneth French's website.

[12]Because NYSE stocks tend to have relatively large capitalizations, this division places more stocks in the low market capitalization (LMC) portfolio than in the high market capitalization (HMC) portfolio.

average returns for each of the six portfolios. The final step is to estimate the excess returns for firm size and financial distress, π_{st} and π_{vt}. The estimated value for π_{st} is the average return on the three "small" stock portfolios (portfolios (4), (5), and (6)) minus the average return on the three "large" stock portfolios ((1), (2), and (3)). Similarly, the estimated value for π_{vt} is the difference between the average return on the two "high" market-to-book portfolios ((1) and (4)) minus the average return on the two "low" market-to-book portfolios ((3) and (6)).

It is also important to correct for infrequent trading when estimating the FF3F model. Accordingly, we also calculate FF3F beta estimates using the following regression:

$$\begin{aligned} r_{it} - r_{ft} &= \alpha_i + \beta_{mi0}(r_{mt} - r_{ft}) + \beta_{mi1}(r_{m,t-1} - r_{f,t-1}) \\ &+ \beta_{si1}\pi_{st} + \beta_{si0}\pi_{s,t-1} + + \beta_{vi1}\pi_{v1t} + \beta_{vi0}\pi_{v,t-1} + \varepsilon_{it} \end{aligned} \tag{8}$$

The sum beta estimates from this model are then defined as: $\hat{\beta}_{ji} = \hat{\beta}_{ji0} + \hat{\beta}_{ji1}$, where j = m for the systematic market risk factor, j = s for the size premium, and j = v for the BV/MV risk premium.

In the second stage of the FF3F methodology, we insert either the betas from equation (7) or the sum beta estimates from equation (8) into equation (3). Also used in this stage are estimates of the long-term average market risk premia π_s and π_v for size and financial distress. As in the case of the market systematic risk factor, the averaging period for the size and financial distress premia is 1926-2000.

The Full-Information Industry Beta (FIB) Approach. The full-information beta (FIB) approach is implemented for both the CAPM and the FF3F models. Because the CAPM can be viewed as a special case of the FF3F model, only the latter model is discussed here.

There are two steps in obtaining beta estimates for any given firm using the FIB approach. The first step is to obtain an estimate of the beta coefficients for each firm in the sample. In the case of the FF3F model, these are the sum beta estimates β_{mi}, β_{si}, and β_{vi} obtained from equation (8). The second step in the FIB method is to conduct cross-sectional regressions with each of the sum beta estimates as dependent variables and industry-participation weights as explanatory variables. The regression, which is conducted with the constant term suppressed, is:

$$\beta_{ji} = \sum_{j=1}^{J} \beta_{fjk}\ \omega_{ik} + v_{ji} \tag{9}$$

where β_{ji} = overall beta estimate of type j for firm i, j = m, s, and v,

β_{fjk} = full-information beta of type j for industry k, j = m, s, and v,

ω_{ik} = industry-participation weight for firm I in industry k, and

v_{ji} = random error term for firm I, equation j.

Following Kaplan and Peterson (1998), we use revenues by industry to calculate ω_{ik}, so that ω_{ij} = revenues of firm i in industry j divided by total revenues of firm i across all industries. We also break down revenues by line of business within the property-liability insurance industry by using net premiums written as the measure of line of business participation in some of our models.

Equation (9) is estimated using two techniques – unweighted least squares (UWLS) and weighted least squares (WLS). In the WLS estimations, the weight for each firm in a specified cross-sectional regression is the ratio of its market capitalization to the total market capitalization of the firms in the sample i.e., the weight is: $S_i / \sum_{i=1}^{N} S_i$, where S_i = market capitalization of firm i and N = the number of firms in the sample.[13] For both the weighted and unweighted cases, we estimate the three Fama-French regressions using the seemingly unrelated regressions (SUR) procedure to improve estimation efficiency by allowing for cross-equation correlations among the regression error terms. The weighted and unweighted FIB regressions for the CAPM are conducted using ordinary least squares.

When UWLS is used to estimate (9) and the corresponding CAPM equation, the β_{fjk} are interpreted as equally-weighted average industry specific betas. When WLS is used, the β_{fjk} represent market value weighted industry betas (Kaplan and Peterson 1998). The equally-weighted results are

---

[13]Kaplan and Peterson (1998) suggest using instrumental variables estimation (IV) rather than weighted least squares (WLS). WLS rather than IV is used here because IV estimation is based on the assumption that the error term in equation (9) is homoskedastic, whereas we consider the assumption of heteroskedasticity to be more appropriate. However, WLS and IV give the same point estimates of the coefficients so the choice only affects the standard errors. Robustness checks revealed that the conclusions of our analysis do not differ materially when IV estimation is used.

useful in obtaining an indication of the betas for the average firm in an industry, whereas the market value weighted (WLS) results are a more useful indicator of the overall cost of capital for an industry.

Consistent with Kaplan and Peterson (1998), we estimate full-information equity betas rather than asset betas. The use of equity betas incorporates the assumption that all firms in an industry have an optimal capital structure and that the firms in the industry are operating at or close to the optimum. Although it would be possible to unlever the estimated equity betas, this approach has the limitation of assuming that all industry segments for a given firm have the same leverage (debt capital to equity capital) ratio. However, it has been established that the appropriate allocation of capital to lines of business within an insurance firm is not likely to lead to equal leverage ratios across lines (Myers and Read 2001).

The Pure Play Approach. As mentioned, the pure play approach involves estimating the cost of capital for firms that have a "pure play" in a given industry, i.e., those firms that specialize in a specific industry rather than diversifying across two or more industries. To estimate pure play costs of capital, we selected a sub-sample from our overall sample of publicly traded property-liability insurers, consisting of all firms that self-report their primary business as property-liability insurance, i.e., that have primary NAICS codes of 524126 or 52413. Robustness checks indicated that similar results would be obtained if we used a rule such as requiring that 75 or 100 percent of revenues come from the property-liability insurance industry. We could not use the pure play approach to estimate the cost of capital by line because there are not enough firms in the sample that specialize in individual lines of business.

4. Empirical Results

We begin this section by discussing summary statistics on the industry participation ratios of property-liability insurers. The overall beta and cost of capital estimation results are then presented, followed by a discussion of betas and costs of capital by line. To illustrate the methodology, we conduct the estimation using three separate line of business categorizations: (1) long-tail versus short-tail lines,

(2) personal versus commercial lines, and(3) automobile insurance versus workers' compensation versus all other lines. The FIB approach could be applied similarly to obtain costs of capital for other lines and line categories.

Summary Statistics

Table 1 presents the industry-participation statistics for publicly traded firms writing property-liability insurance at year-end 2000. The three primary insurance industry categories are presented at the top of the table, followed by non-insurance industries. There are 117 firms in our sample of publicly traded firms that report writing property-liability insurance in 2000. Seventy-five of these firms identify themselves as being primarily property-liability insurers, and 42 firms identify themselves as primarily participating in some other industry. The table shows that firms participating in the property-liability insurance market also are represented in a variety of other industries. The most common industry is "finance excluding insurance," which includes activities such as mutual fund management, financial planning, securities brokerage, and consumer lending. Thirty-four of the firms that list their primary industry as property-liability insurance participate in the finance excluding insurance category, and 22 of the firms in the sample that are not primarily property-liability insurers participate in the finance excluding insurance segment. Only a few firms in the sample earn significant revenues from non-financial industries.

Overall Costs of Capital

In all of the cost of capital estimates presented in this paper, we use as the risk-free rate the 30 day Treasury-bill rate in December 2000. Likewise, as the expected risk-premia for systematic market risk, size, and financial distress, we use the long-run historical (1926 to 2000) market risk premia on NYSE/AMEX/Nasdaq stocks from Kenneth French's web site. We use the same risk-free rate and risk premia for all cost of capital estimates to focus on the impact of the models and the beta coefficients on the cost of capital, holding constant the risk-free rate and market risk premia.

The capital asset pricing model (CAPM) beta and sum-beta estimates for property-liability insurers are summarized in Table 2. The table gives the average beta and sum-beta by market capitalization size quartile for each year of the estimation period. As expected, the sum-beta coefficient estimates are consistently larger than the ordinary beta coefficients. For the sample as a whole, the average beta is 0.677 and the average sum-beta is 0.836. Thus, property-liability insurers on average tend to be characterized by infrequent trading, such that it is important to use sum betas to obtain accurate costs of capital for this industry. Interestingly, the quartile results do not show that large insurers consistently have smaller betas than small insurers, contrary to the usual finding for large and small stocks in general. In part, this is because the size difference between the average large and small property-liability insurers is not as high as for large and small stocks in general, i.e., in year 2000 the average P&L insurer in the largest size quartile is approximately half as large as the average firm in that same quartile.

Table 3 provides the overall beta and sum-beta estimates based on the Fama-French Three Factor (FF3F) method. The beta coefficients for systematic market risk, firm size, and financial distress (BV/MV) are shown by quartile and year of the sample period. On average, the market systematic risk factor has a higher beta coefficient than the BV/MV factor, and the firm size factor has the lowest beta coefficient. For the sample as a whole, the market beta is 0.98, the size beta is 0.386, and the financial distress beta is 0.813. Again, the sum beta estimates are larger than the estimates without the sum beta adjustment, although the differences are not as large as for the CAPM estimates in Table 2.

Comparing our results to the results in Fama and French (1997), we find that our market beta and size beta estimates for property-liability insurers are about the same as the all-industry averages for these two parameters in Fama and French (1997), suggesting that property-liability insurance stocks are about average in terms of their sensitivity to systematic market risk and firm size. However, our financial distress betas, which average 0.813, are <u>substantially</u> larger than the Fama-French all-industry average of 0.02 for this parameter.[14] This result suggests that property-liability stock returns are much more sensitive to financial distress than stocks in general and that financial distress carries a significant cost of capital penalty for property-liability insurers.[15]

Full-Information Costs of Capital

The full-information CAPM beta coefficients for property-liability insurance, life insurance, health insurance, non-insurance finance, and all other industries are shown in Table 4. To obtain the FIB estimates shown in the table, we performed the CAPM full-information regression on all 2-digit NAICS industries using all firms appearing in Compustat that met our sample selection criteria. The dependent variable in the regression is the vector of sum-beta estimates obtained from equation (6). We conducted the FIB estimation separately by year and also conducted a panel data regression including the data from all four years of the sample period in a single regression. Both equally weighted and market value weighted averages are shown in the table. The equally weighted averages provide an indication of the beta for the average insurer, whereas the market value weighted averages provide an indication of the systematic risk sensitivity for the industry as a whole. We focus most of the discussion on the panel data results, but the annual averages are generally quite similar.

Based on the panel regression results, the equally weighted CAPM beta coefficient for the property-liability insurance industry is 0.86 and the value weighted beta is 0.84, i.e., the industry is slightly less risky than stocks in general, which have an average beta coefficient of 1.0. The equally-weighted property-liability industry beta based on the panel estimation model is significantly less than the betas of the health insurance and "all other industries" categories but not significantly different from

[14]Our estimates of the size and financial distress beta for property-liability insurers are also much larger than the Fama-French estimates of these parameters for the insurance industry, 0.09 and 0.06, respectively. However, their definition of the insurance industry is much broader than ours, including life and health insurers as well as property-liability insurers.

[15]Additional evidence of a significant "flight to quality" effect for property-liability insurers is presented in Cummins and Lewis (2003). They analyze a sample of insurers that sustained losses due to the September 11, 2001 World Trade Center terrorist attack and show that stock prices for insurers with high financial ratings rebounded quickly following the attack, whereas stock prices for lower-rated insurers remained depressed.

life insurers or finance excluding insurance. Based on the value-weighted estimates, the property-liability betas are significantly smaller than those of all other industry segments shown in the table. Hence, there is strong evidence that property-liability insurance has lower CAPM systematic risk on average than other industries.

The CAPM costs of capital corresponding to the beta estimates are shown in the last two panels of Table 4. Both the equally weighted and the value weighted estimates suggest that the FIB CAPM cost of equity capital for property-liability insurers is approximately 13.1 percent. Based on the value-weighted estimates, the CAPM cost of capital for property-liability insurers is less than that for life insurers (14.5 percent), health insurers (16.1 percent), financial firms excluding insurers (17.0 percent), and all other industries (14.2 percent).

The FF3F full-information beta estimates and costs of capital for property-liability insurers are shown in Table 5. The estimates in Table 5 incorporate the sum-beta adjustment. The beta coefficients shown in the table are the coefficients of the property-liability industry participation factor in the cross-sectional regressions of the market systematic risk betas (β_{mi}), the small minus big firm capitalization betas (β_{si}), and the high minus low BV/MV betas (β_{vi}), respectively, on the market participation ratios for all Compustat firms meeting our selection criteria. Because the FIB estimates focus only on the property-liability insurance industry component of insurer betas, the numbers differ from those in Table 3, which presents betas for the entire firm rather than specific business lines.

Based on the equally weighted panel regression results for property-liability insurers (panel A of Table 5), the systematic market risk factor is about the same as the financial distress (high-low BV/MV) factor (1.08 and 1.04, respectively). Based on the property-liability insurer value-weighted results (panel C), the financial distress factor is less than the systematic market risk factor (0.686 versus 1.125 based on the panel regression estimates). This suggests that the stock prices of larger insurers are less sensitive to financial distress than those of smaller insurers, as expected if larger firms are more diversified and have better access to capital. The size factors are much smaller than the systematic market risk and financial distress factors for property-liability insurers. The equally weighted panel estimate is around 0.5 and the value-weighted averages are negative (-0.218), suggesting that the small-stock effect does not necessarily apply to the property-liability insurance market as a whole.

Panels B and D of Table 5 show the equally weighted and value weighted FF3F betas for all other industries (excluding property-liability insurance). Based on both the equally weighted and value weighted results, the average financial distress (High-Low Book/Market) beta for property-liability insurers is substantially larger than for average firms in other industries. This provides further evidence that property-liability stock returns are much more sensitive to financial distress than stocks in general and that financial distress carries a significant cost of capital penalty for property-liability firms. The firm size betas are smaller for property-liability insurers than for average firms in other industries, providing further evidence that the size effect is less pronounced for property-liability insurers than for the stock market in general.

The FF3F equally weighted and value weighted cost of capital estimates for property-liability insurers are compared to those for other industries in panels E and F, respectively, of Table 5. Based on the equally weighted results, the cost of capital estimates for property-liability insurance are about the same as for life insurance but significantly larger than for health insurance, finance excluding insurance, and all non-financial industries. Based on the value-weighted results, the cost of capital for property-liability insurance is significantly less than for finance excluding insurance, significantly greater than for all non-financial industries, and not significantly different from life and health insurance.

Perhaps the most important implication of Tables 4 and 5 is that the FF3F costs of capital are substantially larger than the CAPM costs of capital for property-liability insurers. Based on the equally weighted panel estimate results, the FF3F cost of capital for property-liability firms is 21 percent, whereas the CAPM cost of capital is 13.1 percent. Based on the value weighted panel results, the comparison is 18.1 percent for the FF3F method versus 13 percent for the CAPM. The FF3F model leads to higher cost of capital estimates for property-liability insurers than the CAPM for two primary reasons:

(1) the FF3F systematic market risk betas are larger than the comparable CAPM betas, and (2) the FF3F model imposes a positive cost of capital premium for financial distress which is not present under the CAPM. The FF3F size premium has an additional positive effect on the cost of capital based on the equally weighted results but has a slight negative effect based on the value weighted results. Clearly, controlling for factors other than systematic market risk makes a significant difference, suggesting that property-liability insurers relying on the CAPM may be significantly under-estimating the cost of capital.

Table 6 compares the CAPM and FF3F costs of capital for property-liability insurers to the pure play estimates. Recall that the pure play estimates were obtained by restricting the sample to firms that self-report property-liability insurance as their primary industry. Focusing on the annual average results, the pure play costs of capital are slightly smaller than the full-information costs of capital, for all methods shown in the table except the market value weighted FF3F results, where the pure play cost of capital is slightly larger than the FIB result. This general pattern is contrary to the pattern in most other industries, where the pure play estimates tend to exceed the full-information estimates (Kaplan and Peterson 1998), possibly because of the somewhat atypical distribution of firms by size in the property-liability insurance industry. As in the case of the FIB costs of capital, the pure play CAPM estimates are significantly lower than the FF3F estimates, reinforcing the conclusion that focusing on the CAPM alone is likely to underestimate the true cost of capital.

Costs of Capital By Line

In this section, we illustrate the use of the FIB method to estimate the cost of capital by line of property-liability insurance. The results are based on regressions where we replace the Compustat industry-participation variable for property-liability insurers with two or more variables representing line of business distribution within the property-liability insurance industry. For example, if an insurer has its business equally distributed among three lines of insurance and has 75 percent of its total revenues from property-liability insurance, we would replace the 75 percent industry participation ratio with three line of business participation ratios, each equal to 25 percent. As mentioned above, the line of business data are from the NAIC regulatory annual statement files.[16]

The first by-line cost of capital results are presented in Table 7, which shows the full-information CAPM sum-beta cost of capital estimates for short and long-tail lines.[17] Although the equally-weighted beta estimates and costs of capital do not differ significantly between short-tail and long-tail lines, the market-value weighted results are significantly different. For example, the panel regression value-weighted cost of capital for short-tail lines is 17.4 percent whereas the four-year average cost of capital for long-tail lines is 11.1 percent. These results suggest that the short and long-tail costs of capital are about the same for the average insurer but that the long-tail cost of capital is noticeably smaller than the short-tail cost of capital for the property-liability insurance market as a whole.

Table 8 shows the FF3F full-information beta estimates and costs of capital for short-tail and long-tail lines. To conserve space, the beta estimates are provided only for the value-weighted methodology, whereas costs of capital are shown for both the equally-weighted and value-weighted models. Consistent with Table 7, both the equally-weighted and value-weighted results in Table 8 show that the cost of capital is lower for long-tail lines than for short-tail lines. The panel regression short-tail and long-tail costs of capital are 19.4 percent versus 17.3 percent based on the equally weighted results and 25.6 percent versus 13.1 percent based on the value weighted results, although the difference is

---

[16]We calculate the property-liability insurance line-of-business participation weights by multiplying the percentage of the firm's statutory premiums in a particular line of insurance by the firm's overall proportion of net sales in the property-liability insurance industry calculated using the consolidated GAAP revenue data reported on Compustat. An alternative way to calculate the property-liability line-of-business participation weights would have been to divide the firm's total statutory premiums in the particular line of insurance by the total net sales for the firm as reported on Compustat. We do not use this latter method owing to the differences that exist between GAAP and Statutory accounting rules and because the NAIC data files only contain information on the insurer's domestic U.S. business while the GAAP consolidated data contains net sales of the insurer's domestic and foreign subsidiaries. Our preferred method makes the assumption that the insurer's foreign property-liability insurance business is divided among the various property-liability lines of insurance in proportions similar to its domestic business.

[17]Short-tail lines of insurance includes property coverages (e.g. fire, allied lines, homeowners multi-peril, automobile physical damage etc.), all accident and health coverages and all financial guaranty businesses (fidelity, surety, mortgage guaranty, etc.). Long-tail business includes all liability insurance coverages (other liability, products liability, personal and commercial automobile liability) and all reinsurance.

statistically significant only in the value-weighted case. Therefore, it appears that more highly capitalized insurers can write long-tail lines of insurance at much lower costs of capital than smaller insurers, perhaps because of their better diversification, better access to capital, and greater ability to withstand large loss shocks.

The finding that long-tail lines have lower costs of capital than short-tail lines seems to be contrary to the conventional wisdom in the insurance industry that long-tail lines are generally riskier than short-tail lines.[18] One possible explanation for this finding is that the market value of liabilities in long-tail lines is more sensitive to changes in interest rates than is the case for short-tail lines. Because the market values of assets and liabilities tend to move in the same direction in response to interest rate changes, insurers focusing on long-tail lines thus tend to have a "natural hedge" against interest rate risk, which is much less evident in the short-tail lines, possibly leading to lower costs of capital in the long-tail lines. Short-tail lines are also more susceptible to catastrophic losses from events such as hurricanes and earthquakes, providing another possible explanation for the cost of capital difference.

The next set of cost of capital decompositions is based on the personal lines versus the commercial lines of property-liability insurance.[19] The full-information CAPM beta estimates with the sum beta adjustment for personal and commercial lines are shown in Table 9. The panel regression equally-weighted results show a cost of capital of 13.8 percent for the personal lines and 12.8 percent for the commercial lines, whereas the ordering is reversed in the value weighted estimates (11.7 percent for personal lines and 13.6 percent for commercial lines). The difference is statistically significant for the value weighted but not for the equally weighted results. Thus, for the average insurer, the cost of

---

[18]Our results are not necessarily inconsistent with the finding by Cummins and Lamm-Tennant (1994) that the cost of capital is higher for commercial long-tail lines than for all other lines. As reported below, our results show a higher cost of capital for commercial lines than for personal lines, so it is difficult to say whether the Cummins-Lamm-Tennant finding is driven by commercial lines or long-tail lines (they include only a single business mix variable representing commercial long-tail lines). Additionally, comparison between our results and theirs may not be meaningful, given that their sample period ended in 1989.

[19]Personal lines of insurance includes homeowners, farmowners, earthquake, personal automobile liability and automobile physical damage. All other lines of insurance are considered commercial lines.

capital is slightly higher for personal lines, but for the market as a whole the cost of capital is higher for commercial lines. This may indicate that the types of commercial business written by larger insurers (e.g., national and multi-national accounts) are more risky than those written by smaller insurers, which tend to focus on local or regional risks. In addition, it may reflect the superior ability of larger insurers to cover catastrophic personal lines property risks because of their better capitalization.

The full-information betas and costs of capital for personal versus commercial lines based on the Fama-French methodology are shown in Table 10. As with the CAPM results, the equally weighted cost of capital estimates in Table 10 imply that the cost of capital for the average insurer is slightly larger for personal lines than for commercial lines (23.1 versus 19.5 percent based on the panel regression). However, the value weighted results show the opposite relationship, costs of capital of 16.3 percent versus 18.8 percent based on the panel regression. These results thus provide additional evidence to suggest that the commercial lines have a higher cost of capital than the personal lines for the market as a whole but not for insurers on average.

The final set of cost of capital decompositions is based on the subdivision of property-liability insurance into automobile insurance, workers' compensation, and all other property-liability lines combined. This line breakdown was chosen to focus on the two most heavily price-regulated lines – automobile and workers' compensation insurance.[20] The CAPM sum-beta results are reported in Table 11. Based on the equally weighted panel estimate results, the cost of capital for automobile insurance is slightly higher than for workers' compensation (13.6 percent versus 13.3 percent), although this difference is not statistically significant. Based on the value-weighted results, the cost of capital for automobile insurance is less than for workers' compensation insurance – 11.3 percent versus 13.4

[20]Automobile insurance includes premiums written in personal and commercial automobile liability insurance and in automobile physical damage. Further decomposition of the automobile insurance line of business showed that personal automobile liability (the most heavily regulated automobile insurance line) had higher costs of capital than automobile insurance in the aggregate (results available from the authors). Data to decompose automobile physical damage into personal and commercial components are not available in our data source.

percent, although again the difference is not statistically significant. The cost of capital for all other property-liability (P&L) lines of business is not statistically different from the costs of capital of automobile and workers' compensation insurance based on the equally weighted results, but automobile insurance has a significantly lower cost of capital than all other lines based on the value weighted results. The conclusions to be drawn from Table 11 are that the CAPM costs of capital are about the same for automobile and workers' compensation insurance but that the value weighted (market wide) cost of capital for automobile insurance is significantly lower than for all other lines combined.

Another important inference from Table 11 is that the market wide (value weighted) cost of capital for automobile insurance is significantly lower than the cost of capital for the average insurer for this line. This result illustrates one of the hazards of insurance price regulation, which tends to be based on industry-wide costs of capital rather than costs of capital by firm. As Table 11 indicates, basing prices on industry-wide results could lead to significant pricing errors for many firms in the industry.

The FF3F full-information beta and cost of capital estimates for automobile insurance, workers' compensation, and all other lines are presented in Table 12. As in Tables 8 and 10 only the value-weighted beta estimates are shown in order to conserve space. However, both equally-weighted and value-weighted costs of capital are shown in the table. The results in Table 12 show that the FF3F method again leads to cost of capital estimates that are significantly higher than the CAPM cost of capital estimates (Table 11). For example, the equally-weighted panel estimate costs of capital are 22.3 percent for automobile insurance, 19.6 percent for workers' compensation, and 19.8 percent for all other lines, compared to 13.6 percent for auto, 13.3 percent for workers' compensation, and 12.7 percent for all other lines based on the CAPM. The value-weighted FF3F estimates are also higher than those for the CAPM.

The results in Table 12 also have important implications for price regulation. The results suggest that failure to recognize sources of risk other than the CAPM market systematic risk factor could lead to significant underpricing in regulated lines. In addition ,the results reinforce the conclusion based on

Table 11 that the industry-wide cost of capital is significantly lower than the average-firm cost of capital for automobile insurance. Thus, basing prices on industry-wide costs of capital is likely to be value-destroying for the average firm in the industry.

5. Conclusions

This paper investigates the estimation of the cost of equity capital for property-liability insurers using a relatively new methodology, the full-information beta approach. The method is designed to obtain the cost of capital for a division or line of business of a firm, where the divisions (business lines) of the firm are not publicly traded. The procedure is to obtain the beta coefficients for a sample of firms and then to regress the betas cross-sectionally against variables measuring the firm's business composition across industries. The business composition variables used in this study are the ratios of the revenues coming from each industry divided by total revenues from all industries. The estimated regression coefficients are interpreted as full-information betas.

We obtain beta coefficient estimates using two principal cost of capital models in order to implement the full information beta approach – the capital asset pricing model (CAPM) and the Fama-French three-factor model (FF3F). The CAPM includes a single risk factor representing the firm's exposure to systematic market risk. The FF3F model augments the systematic risk factor by adding factors for firm size (total market capitalization) and the financial distress of the firm, proxied by the ratio of the book value (BV) of equity to the market value (MV) of equity. Based on prior empirical research, firm size is expected to be inversely related to the cost of capital, and the BV/MV ratio is expected to be positively related to the cost of capital. In estimating the beta coefficients for the CAPM and the FF3F method, we utilize the sum-beta procedure designed to adjust for the problem of infrequent trading. This adjustment is especially important for the property-liability insurance industry because many insurance stocks are characterized by infrequent trading.

To estimate the full-information betas for the property-liability insurance industry, we utilize a sample consisting of all Compustat firms meeting our sampling criteria, for the estimation period 1997-

2000. The sample includes 172 publicly traded firms writing property-liability insurance. Industry participation variables are included for all two-digit industries defined by the North American Industry Classification System (NAICS). The coefficient of the industry-participation ratio for a particular industry is then interpreted as the full-information beta coefficient for that industry. For the CAPM, only one FIB regression is conducted, with the market systemic risk factor (beta) as the dependent variable. For the FF3F method, three FIB regressions are estimated, one for each of the three factors in the Fama-French model.

In the first set of full-information beta regressions considered in the paper, we estimate the full-information betas for the entire property-liability insurance industry using only Compustat data to obtain the industry-participation ratios. In the second set of regressions, we utilize data from the National Association of Insurance Commissioners to break down the revenues of property-liability insurers by line of insurance. We estimate full information betas for three different insurance-line groupings – (1) long-tail versus short-tail lines, (2) personal versus commercial lines; and (3) automobile insurance, workers' compensation insurance, and all other lines

The primary conclusions of the paper are the following:

1. It is important to use the sum beta technique to control for infrequent trading when estimating betas for the property-liability insurance industry. Failure to adjust for this problem is likely to lead to under-estimation of the cost of capital.

2. The cost of capital estimates from the FF3F method are generally higher than the estimates based on the CAPM. Hence, failure to adjust for firm size and financial distress could lead to significant under-estimation of the cost of capital.

3. The cost of capital varies significantly by line of insurance and also varies between large and small insurers. Thus, it is important to use firm and line specific costs of capital in applications such as project selection and capital allocation.

4. Value-weighted estimates of the cost of capital often differ significantly from equally weighted estimates. This suggests that basing price regulation on industry-wide results rather than costs of capital by firm may lead to significant pricing errors for many firms in the industry.

In general, we believe that the full-information beta approach provides a reliable method for estimating costs of capital by line for property-liability insurers. The method is likely to obtain the most reliable results if it is used with the sum beta adjustment and the FF3F method. The CAPM cost of capital with sum beta adjustment also may be useful at least as a reasonability check on the FF3Fmethod.

Full-information betas can be used by insurers in a variety of contexts, including the allocation of capital by line of business, estimation of risk adjusted returns on capital (RAROC), insurance pricing, and decision making about entering or exiting lines of business. Full-information costs of capital also could be used to evaluate potential merger and acquisition transactions. Another important advantage is that the full-information model can be used to estimate the cost of capital for insurers that do not have traded equity, including mutuals, reciprocals, and untraded stock insurers owned by publicly traded insurers, financial services firms, or conglomerates.

References

Brealey, Richard A. and Stewart C. Myers, 2002, *Principles of Corporate Finance*, 7th ed. Boston: McGraw-Hill/Irwin.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Cochrane, John H, 1999, "New Facts in Finance," *Economic Perspectives* 23(3): 36-58.

Cox, Larry A. and Gary L. Griepentrog, 1988, "The Pure-Play Cost of Equity for Insurance Divisions," *Journal of Risk and Insurance* 55: 442-452.

Cummins, J. David, 1990, "Multi-Period Discounted Cash Flow Ratemaking Models in Property-Liability Insurance," *Journal of Risk and Insurance* 57: 79-109.

Cummins, J. David and Scott E. Harrington, eds., 1987, *Fair Rate of Return in Property-Liability Insurance*. Hingham, MA: Kluwer Academic Publishers.

Cummins, J David and Scott E. Harrington, Scott, 1985, "Property-Liability Insurance Rate Regulation: Estimation of Underwriting Betas Using Quarterly Profit Data," *Journal of Risk and Insurance* 52: 16-43.

Cummins, J. David and Joan Lamm-Tennant, 1994, "Capital Structure and the Cost of Equity Capital in Property-Liability Insurance," *Insurance: Mathematics and Economics* 15: 187-201.

Cummins, J. David and Christopher M. Lewis, 2003, "Catastrophic Events, Parameter Uncertainty, and the Breakdown of Implicit Long-Term Contracting: The Case of Terrorism Insurance," *Journal of Risk and Uncertainty*, forthcoming.

Cummins, J. David and Richard D. Phillips, 2001, "Financial Pricing of Property-Liability Insurance," in Georges Dionne, ed., *Handbook of Insurance Economics* (Boston, MA: Kluwer Academic Publishers).

Derrig, Richard A., 1990, "The Development of Property-Liability Insurance Pricing Models in the United States 1969-1989," Proceedings of the 1st AFIR International Colloquium, Paris, France.

Derrig, Richard A. ,and Elisha Orr, 2003, "Equity Risk Premium: Expectations Great and Small," presented at the 2003 Bowles Symposium, Georgia State University, Atlanta, GA.

Dimson, Elroy, 1979, "Risk Measurement When Shares Are Subject To Infrequent Trading," *Journal of Financial Economics* 7: 197-226.

Ehrhardt, Michael C. and Yatin N. Bhagwat, 1991, "A Full-Information Approach for Estimating Divisional Betas," *Financial Management* 20: 60-69.

Elton, Edward J. and Martin J. Gruber, 1991, *Modern Portfolio Theory and Investment Analysis*, 4th edition. New York: John Wiley & Sons.

Fama, Eugene F. and Kenneth R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47: 427-465.

Fama, Eugene F. and Kenneth R. French, 1993, "Common Risk Factors In the Returns on Stocks and Bonds," *Journal of Financial Economics* 39: 3-56.

Fama, Eugene F. and Kenneth R. French, 1997, "Industry Costs of Equity," *Journal of Financial Economics* 43: 153-193.

Fuller, Russell J. and Halbert S. Kerr, 1981, "Estimating the Divisional Cost of Capital: An Analysis of the Pure-Play Technique," *Journal of Finance* 36: 997-1009.

Girard, Luke N., 2002, "An Approach to Fair Valuation of Insurance Liabilities Using the Firm's Cost of Capital," *North American Actuarial Journal* 6:18-41.

Graham, John R. and Campbell R. Harvey, 2001, "The Theory and Practice of Corporate Finance: Evidence from the Field," *Journal of Financial Economics* 60: 187-243.

Ibbotson Associates, 2001, *Stocks, Bonds, Bills, and Inflation: Valuation Edition (2001 Yearbook)* Chicago.

Ibbotson, Roger G., Paul D. Kaplan, and James D. Peterson, 1997, "Estimates of Small-Stock Betas Are Much Too Low," *Journal of Portfolio Management*

Ingersoll, Jonathan E., Jr., 1987, Theory of Financial Decision Making . Totowa, NJ: Rowman & Littlefield.

Kaplan, Paul D. and James D. Peterson, 1998, "Full-Information Industry Betas," *Financial Management* 27: 85-93.

Kolbe, Lawrence, James Read, and Gregory R. Hall, 1984, *Cost of Capital: Estimating the Rate of Return for Public Utilities.* Cambridge, MA: MIT Press.

Krueger, Mark K. and Charles M. Linke, 1994, "A Spanning Approach for Estimating Divisional Cost of Capital," *Financial Management* 23: 64-70.

Lee, Alice C. and J. David Cummins, 1998, "Alternative Models for Estimating the Cost of Equity Capital for Property/Casualty Insurers," *Review of Quantitative Finance and Accounting* 10: 235-267.

Myers, Stewart C and James A. Read, Jr., 2001, "Capital Allocation for Insurance Companies," *Journal of Risk & Insurance* 68: 545-580.

Panjer, Harry H., 1998, *Financial Economics With Applications to Investments, Insurance, and Pensions*. Schaumburg, IL: The Actuarial Foundation.

Roll, Richard and Stephen A. Ross, 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance* 35: 1073-1103.

Ross, Stephen A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13(2): 383-402.

Schink, George R. and Richard S. Bower, 1994, "Application of the Fama-French Model to Utility Stocks," *Financial Markets, Institutions and Instruments* 3: 74-96.

Scholes, Myron and Joseph Williams, 1977, "Estimating Beta from Nonsynchronous Data," *Journal of Financial Economics* 5(3): 309-327.

Taylor, Greg, 1994, "Fair Premium Rating Methods and the Relations between Them," *Journal of Risk & Insurance* 61: 592-615.

Thompson, Howard E., 1991, *Regulatory Finance* (Boston: Kluwer Academic Publishers).

Wang, Kevin Q., 2003, "Asset Pricing with Conditioning Information: A New Test," *Journal of Finance* 58: 161-196.

Wei, K.C. John, 1988, "An Asset-Pricing Theory Unifying the CAPM and APT," *Journal of Finance* 43: 881-892.

Table 1

Sources of Revenue Across Industries for Firms Writing Property-Liability Insurance: 2000

Table displays the total revenue across industries for all firms writing property-liability insurance during 2000. Industry classifications were derived using the North American Industry Classification System (NAICS). The firm's self-reported primary NAICS code was used to classify the firms as either predominantly P&L insurers or as firms whose primary business line was something else. Revenue amounts are reported in $millions. The number of firms with positive revenues in a given industry segment are reported in parantheses. Source: COMPUSTAT Business Information File.

2 - Digit NAICS	Industry Description	Primarily Non-P&L Insurer		Primarily P&L Insurer	
-	Property/Liability Ins. and Reinsurance*	$ 71,639.26	(42)	$ 134,218.70	(75)
-	Health Insurance*	4,050.54	(5)	9,283.92	(2)
-	Life Insurance*	93,430.59	(10)	16,664.25	(17)
72	Accommodation and Food Services	-	-	351.06	(1)
71	Arts, Entertainment, and Recreation	-	-	32.68	(1)
62	Health Care and Social Assistance	4.96	(1)	-	-
61	Education Services	1,856.83	(1)	-	-
56	Administrative Support, Waste Management and Remediation	626.64	(2)	-	-
54	Professional, Scientific, and Technical Services	488.09	(3)	350.29	(2)
53	Real Estate and Rental and Leasing	5,078.98	(4)	8.20	(1)
52	Finance Excluding Insurance*	26,456.14	(22)	5,518.24	(34)
51	Information	326.12	(4)	3.55	(1)
48	Transportation and Warehousing	99.50	(1)	31.52	(1)
45	Retail Trade 2	332.43	(1)	-	-
44	Retail Trade 1	1,402.19	(1)	-	-
42	Wholesale Trade	-	-	378.60	(3)
33	Manufacturing - Heavy Ind., Machinery, Electronic & Computer	2,750.82	(2)	-	(1)
32	Manufacturing - Light Commercial Products	1,607.42	(1)	63.98	(1)
31	Manufacturing - Consumer Items	803.50	(1)	4,302.81	(2)
23	Construction	13.15	(1)	155.68	(2)
22	Mining	-	-	1,054.52	(2)

Note: Property-liability insurers and reinsurers have NAICS codes 524126 and 52413, respectively. Life insurers have NAICS code 524113. Health insurers have NAICS code 524114. All other NAICS codes under the "Finance and Insurance" heading (2 Digit NAICS code 52) are classified as "Finance Excluding Insurance."

Table 2

Summary Statistics for CAPM Regressions on Property-Liability Insurers

Table shows average CAPM beta for firms which self-identify as property-liability insurers by listing their overall NAICS code as 524126 or 52413. Both beta and sum beta regressions are conducted for each firm (equations (5) and (6)). The beta regression is:

$$(R_i - R_f) = a + \beta(R_m - R_f) + e$$

where R_i is the return on firm i, R_f is the one-month Treasury bill rate observed at the beginning of the month, and R_m is the value-weighted market return on all NYSE, AMEX, and NASDAQ stocks. The "Sum β" model adjusts for non-synchronous trading by adding to the regression the excess market return variable lagged one time period. The reported statistic equals the sum of the contemporaneous and lagged beta estimates. The data period for each year end on June 30. Estimates are calculated using the previous 60-months of returns.

Year	Market Cap Quartile	No. P&L Insurers	Average β	Average Sum β
1997	Small	21	0.646	0.893
	2	21	0.861	1.144
	3	21	0.709	0.809
	Big	22	0.820	0.932
	Total	85	0.760	0.944
1998	Small	18	0.632	0.926
	2	19	0.687	0.908
	3	19	0.652	0.811
	Big	19	0.917	0.999
	Total	75	0.723	0.911
1999	Small	19	0.570	0.812
	2	19	0.616	0.677
	3	19	0.642	0.736
	Big	19	0.690	0.746
	Total	76	0.629	0.743
2000	Small	18	0.316	0.631
	2	18	0.654	0.763
	3	18	0.642	0.696
	Big	19	0.712	0.817
	Total	73	0.583	0.728
Grand Total		309	0.677	0.836

Table 7

Full Information CAPM Beta Estimates with Sum Beta Adjustment: Short-Tail Lines vs. Long-Tail Lines

Table displays full information CAPM beta estimates for by-line of property-liability Insurance. Short-tail lines of insurance include all property coverages (fire, allied, etc.), financial and mortgage guaranty, fidelity and surety, etc. Long-tail lines of insurance includes all liability coverages (products liability, other liability, etc.) as well as reinsurance, international etc. See footnote 17 for a complete categorization of all lines. The full-informaton beta comes from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is the equity beta estimated using equation (6) for firm i, $\beta full_j$ is the estimated full-information beta for industry (line of business) j, wij is the percent of firm i's net sales in industry (line of business) j. The regression is estimated using OLS and via weighted least squares to obtain market-capitalization weighted industry (line of business) full-information betas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated beta greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated separately for each calendar year and as a pooled regression across all four years. The risk-free rate of interest is the 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical market risk premium as of December 2000 was 8.49 percent.

	1997	1998	1999	2000	Average	Panel Estimate
Beta (Equally Weighted)						
Short-tail Lines	0.945	0.953	0.882	0.696	0.869	0.876
	(0.232)	(0.223)	(0.203)	(0.221)		(0.111)
Long-tail Lines	0.926	0.895	0.720	0.806	0.837	0.842
	(0.151)	(0.154)	(0.140)	(0.150)		(0.075)
F-Test: $\beta_{Short\text{-}tail} = \beta_{Long\text{-}tail}$	0.000	0.040	0.350	0.130		0.050
Beta (Market Value Weighted)						
Short-tail Lines	1.656	1.400	1.110	1.075	1.310	1.352
	(0.149)	(0.119)	(0.161)	(0.212)		(0.078)
Long-tail Lines	0.525	0.587	0.589	0.925	0.657	0.609
	(0.075)	(0.068)	(0.094)	(0.119)		(0.043)
F-Test: $\beta_{Short\text{-}tail} = \beta_{Long\text{-}tail}$	35.69 ***	28.11 ***	6.08 **	0.290		54.24 ***
Cost of Equity Capital (Equally Weighted)						
Short-tail Lines	13.9%	14.0%	13.4%	11.8%	13.3%	13.3%
Long-tail Lines	13.7%	13.5%	12.0%	12.7%	13.0%	13.0%
Cost of Equity Capital (Market Value Weighted)						
Short-tail Lines	19.9%	17.8%	15.3%	15.0%	17.0%	17.4%
Long-tail Lines	10.3%	10.9%	10.9%	13.7%	11.5%	11.1%

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 4

Full Information CAPM Beta Estimates with Sum Beta Adjustment By Industry Category

Table displays full information CAPM beta estimates for property-liability insurance, life insurance, health hnsurance, finance excluding insurance, and all other industries. The full-informaton beta is estimated from the following cross-sectional regression

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is the equity beta estimated using equation (6) for firm i, $\beta full_j$ is the estimated full-information beta for industry j, W_{ij} is the percent of firm i's net sales in industry j. The regression is estimated by OLS (Equally weighted) and via wieghted least squares (Market Weighted). The latter is used so we can obtain market-capitalization weighted industry full-information betas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated beta greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated separately for each calendar year and as a pooled regression across all four years. The risk-free rate of interest used to estimate the cost of equity capital was the published 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical market risk premiun as of December 2000 was 8.49 percent (Ibottson, 2001).

	1997	1998	1999	2000	Average	Panel Estimate
Beta (Equally Weighted)						
Property-Liability Insurance	0.939	0.931	0.769	0.759	0.849	0.856
	(0.109)	(0.108)	(0.099)	(0.103)		(0.053)
Life Insurance	0.974	0.952	0.821	0.785	0.883	0.897
	(0.182)	(0.198)	(0.193)	(0.190)		(0.095)
Health Insurance	0.722	1.023	1.584	1.160	1.122	1.096
	(0.203)	(0.229)	(0.190)	(0.214)		(0.105)
Finance (Excluding Insurance)	0.915	0.822	0.868	0.782	0.847	0.847
	(0.060)	(0.057)	(0.050)	(0.190)		(0.027)
Average All Non-Financial Industries	1.108	1.082	1.099	1.057	1.087	1.091
F-Test: $\beta_{P\&L} = \beta_{Life}$	0.050	0.070	0.030	0.010		0.200
F-Test: $\beta_{P\&L} = \beta_{Health}$	0.750	0.230	13.80 ***	2.700		4.40 **
F-Test: $\beta_{P\&L} = \beta_{Finance(Excluding\ Ins)}$	0.010	0.310	0.580	0.010		0.000
F-Test: $\beta_{P\&L} = \beta_{Average\ All\ Non\text{-}Financial\ Industries}$	2.450	2.530	9.24 ***	7.06 ***		18.65 ***
Beta (Market Value Weighted)						
Property-Liability Insurance	0.847	0.858	0.764	0.964	0.858	0.843
	(0.054)	(0.050)	(0.067)	(0.085)		(0.031)
Life Insurance	1.114	0.971	0.880	1.057	1.006	1.012
	(0.097)	(0.088)	(0.104)	(0.136)		(0.059)
Health Insurance	1.007	1.434	1.324	1.150	1.229	1.208
	(0.103)	(0.112)	(0.113)	(0.141)		(0.053)
Finance (Excluding Insurance)	1.327	1.199	1.365	1.330	1.305	1.309
	(0.036)	(0.031)	(0.029)	(0.032)		(0.016)
Average All Non-Financial Industries	1.026	0.983	0.968	0.987	0.991	0.986
F-Test: $\beta_{P\&L} = \beta_{Life}$	5.97 **	2.78 *	0.940	0.260		9.23 ***
F-Test: $\beta_{P\&L} = \beta_{Health}$	2.490	24.79 ***	18.30 ***	1.150		32.88 ***
F-Test: $\beta_{P\&L} = \beta_{Finance(Excluding\ Ins)}$	55.95 ***	40.42 ***	67.70 ***	15.76 ***		184.75 ***
F-Test: $\beta_{P\&L} = \beta_{Average\ All\ Non\text{-}Financial\ Industries}$	10.09 ***	8.86 ***	8.16 ***	0.020		22.22 ***
Cost of Equity Capital (Equally Weighted)						
Property-Liability Insurance	13.9%	13.8%	12.4%	12.3%	13.1%	13.1%
Life Insurance	14.2%	14.0%	12.9%	12.5%	13.4%	13.5%
Health Insurance	12.0%	14.6%	19.3%	15.7%	15.4%	15.2%
Finance (Excluding Insurance)	13.6%	12.9%	13.2%	12.5%	13.1%	13.1%
Average All Non-Financial Industries	15.3%	15.1%	15.2%	14.9%	15.1%	15.1%
Cost of Equity Capital (Market Value Weighted)						
Property-Liability Insurance	13.1%	13.2%	12.4%	14.1%	13.2%	13.0%
Life Insurance	15.3%	14.1%	13.4%	14.8%	14.4%	14.5%
Health Insurance	14.4%	18.1%	17.1%	15.6%	16.3%	16.1%
Finance (Excluding Insurance)	17.1%	16.1%	17.5%	17.2%	17.0%	17.0%
Average All Non-Financial Industries	14.6%	14.2%	14.1%	14.3%	14.3%	14.2%

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 5

Full Information Fama-French 3-Factor Estimates with Sum Beta Adjustment: Property-Liability Insurance

Table displays full information beta estimates for the Fama-French 3 Factor Model for the property-liability insurance industry. The full-information beta for each factor is the estimated from the following cross-sectional regression

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_i$$

where β_i is either the excess market coefficient, the SMB coefficient, or the HML coefficient estimated using equation (8) for firm i, $\beta full_j$ is the estimated full-information coefficient estimate for industry j, Wij is the percent of firm i's net sales in industry j. The regression is estimated by OLS (Equally Weighted) and via weighted least squares (Market Weighted). The latter is used so we can obtain market-capitalization weighted industry full-information betas. The weight is equal to the market capitalization of firm I relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks with an estimated factor coefficient greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated each calendar year and as a pooled regression across all four years. The risk-free rate of interest used to estimate the cost of equity capital was the published 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical premium for the excess market return,the SMB factor, and the HML factor as of December 2000 was 8.49 percent, 2.21 percent, and 4.63 percent respectively.

	1997	1998	1999	2000	Average	Panel Estimate
Panel A: Property-Liability Equally Weighted						
CAPM Beta Factor β_m	1.113	1.099	0.958	1.143	1.078	1.080
	(0.106)	(0.106)	(0.103)	(0.109)		(0.053)
Small - Big Capitalization Factor β_s	0.353	0.586	0.618	0.472	0.507	0.501
	(0.136)	(0.133)	(0.124)	(0.123)		(0.065)
High - Low Book/Market Factor β_v	0.813	1.066	1.026	1.311	1.054	1.040
	(0.139)	(0.169)	(0.153)	(0.155)		(0.077)
Panel B: Average of All Other Industries Equally Weighted[1]						
CAPM Beta Factor β_m	0.873	0.907	0.880	1.083	0.936	0.937
Small - Big Capitalization Factor β_s	0.972	0.940	0.909	0.663	0.871	0.866
High - Low Book/Market Factor β_v	0.185	0.184	0.209	0.585	0.291	0.284
Panel C: Property-Liability Market Value Weighted						
CAPM Beta Factor β_m	1.021	1.131	1.148	1.355	1.164	1.125
	(0.054)	(0.053)	(0.068)	(0.082)		(0.031)
Small - Big Capitalization Factor β_s	-0.387	-0.211	-0.015	-0.126	-0.185	-0.218
	(0.080)	(0.080)	(0.077)	(0.097)		(0.042)
High - Low Book/Market Factor β_v	0.230	0.720	1.020	1.150	0.780	0.686
	(0.085)	(0.106)	(0.102)	(0.139)		(0.054)
Panel D: Average of All Other Industries Market Value Weighted[1]						
CAPM Beta Factor β_m	0.941	0.974	1.010	1.193	1.030	1.003
Small - Big Capitalization Factor β_s	0.337	0.240	0.155	-0.026	0.177	0.186
High - Low Book/Market Factor β_v	0.051	0.101	0.195	0.448	0.199	0.148
Panel E: Industry Cost of Equity Capital Equally Weighted						
Property-Liability Insurance	19.9%	21.4%	20.1%	22.7%	21.0%	21.0%
Life Insurance	20.7%	19.9%	20.0%	21.3%	20.5%	20.5%
Health Insurance	6.3%	13.3%	19.0%	22.0%	15.1%	14.3%
Finance Excluding Insurance	19.4%	18.0%	16.0%	16.7%	17.5%	17.6%
Average All Non-Financial Industries	16.3%	16.5%	16.3%	19.3%	17.1%	17.1%
F-Test: $r_{P\&L} = r_{Life}$	0.100	0.230	0.000	0.240		0.100
F-Test: $r_{P\&L} = r_{Health}$	22.51 ***	5.35 **	0.200	0.040		18.46 ***
F-Test: $\beta_{P\&L} = \beta_{Finance(Excluding\ Ins)}$	0.090	4.02 **	8.18 ***	15.22 ***		19.06 ***
F-Test: $r_{P\&L} = r_{Average\ All\ Non\text{-}Financial\ Industries}$	6.18 **	10.34 ***	9.87 ***	6.46 **		31.11 ***
Panel F: Industry Cost of Equity Capital Market Value Weighted						
Property-Liability Insurance	14.8%	18.4%	20.3%	22.4%	19.0%	18.1%
Life Insurance	23.7%	19.9%	13.7%	13.7%	17.8%	18.8%
Health Insurance	9.9%	17.6%	20.6%	23.5%	17.9%	16.9%
Finance Excluding Insurance	22.3%	23.0%	20.2%	19.5%	21.2%	21.1%
Average All Non-Financial Industries	14.8%	15.1%	15.7%	18.0%	15.9%	15.5%
F-Test: $r_{P\&L} = r_{Life}$	27.89 ***	0.610	12.98 ***	13.11 ***		0.530
F-Test: $r_{P\&L} = r_{Health}$	9.06 ***	0.110	0.030	0.260		1.470
F-Test: $\beta_{P\&L} = \beta_{Finance(Excluding\ Ins)}$	66.88 ***	17.85 ***	0.020	6.29 **		33.01 ***
F-Test: $r_{P\&L} = r_{Average\ All\ Non\text{-}Financial\ Industries}$	0.280	14.74 ***	25.93 ***	12.81 ***		38.29 ***

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 6
Property-Liability Insurer Costs of Capital:
Full-Information Betas Versus the Pure Play Methodology

Table displays the cost of capital for the property-liability industry calculated using the full information beta methodology ver: the average cost of capital for firms which self-report their primary business being property-liability insurance (primary NAIC codes 524126 and 52413). The risk-free rate of interest used to estimate the cost of equity capital was the published 30 da T-bill rate in December 2000, 5.88 percent. The long-run historical premium for the excess market return, the SMB factor, a the HML factor as of December 2000 was 8.49 percent, 2.21 percent, and 4.63 percent, respectively

	1997	1998	1999	2000	Average	Panel Estimate
Full Information Beta Methodology Estimates						
Equally Weighted						
CAPM Beta	12.6%	12.2%	11.5%	11.0%	11.8%	11.9%
CAPM Sum Beta	13.9%	13.8%	12.4%	12.3%	13.1%	13.1%
Fama-French	19.2%	19.5%	19.1%	21.3%	19.8%	19.7%
Fama-French Sum Beta	19.9%	21.4%	20.1%	22.7%	21.0%	21.0%
Market Value Weighted						
CAPM Beta	13.8%	13.9%	12.4%	12.8%	13.2%	13.4%
CAPM Sum Beta	13.1%	13.2%	12.4%	14.1%	13.2%	13.0%
Fama-French	17.9%	19.2%	20.4%	22.7%	20.0%	19.5%
Fama-French Sum Beta	14.8%	18.4%	20.3%	22.4%	19.0%	18.1%
Average of Pure Play Estimates						
CAPM Beta	12.3%	12.0%	11.2%	10.8%	11.6%	-
CAPM Sum Beta	13.9%	13.6%	12.2%	12.1%	12.9%	-
Fama-French	18.4%	18.9%	17.8%	20.3%	18.8%	-
Fama-French Sum Beta	19.4%	20.8%	18.9%	21.8%	20.2%	-

Table 7

Full Information CAPM Beta Estimates with Sum Beta Adjustment: Short-Tail Lines vs. Long-Tail Lines

Table displays full information CAPM beta estimates for by-line of property-liability Insurance. Short-tail lines of insurance include all property coverages (fire, allied, etc.), financial and mortgage guaranty, fidelity and surety, etc. Long-tail lines of insurance includes all liability coverages (products liability, other liability, etc.) as well as reinsurance, international etc. See footnote 17 for a complete categorization of all lines. The full-informaton beta comes from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is the equity beta estimated using equation (6) for firm i, $\beta full_j$ is the estimated full-information beta for industry (line of business) j, wij is the percent of firm i's net sales in industry (line of business) j. The regression is estimated using OLS and via weighted least squares to obtain market-capitalization weighted industry (line of business) full-information betas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated beta greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated separately for each calendar year and as a pooled regression across all four years. The risk-free rate of interest is the 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical market risk premium as of December 2000 was 8.49 percent.

	1997	1998	1999	2000	Average	Panel Estimate
Beta (Equally Weighted)						
Short-tail Lines	0.945	0.953	0.882	0.696	0.869	0.876
	(0.232)	(0.223)	(0.203)	(0.221)		(0.111)
Long-tail Lines	0.926	0.895	0.720	0.806	0.837	0.842
	(0.151)	(0.154)	(0.140)	(0.150)		(0.075)
F-Test: $\beta_{Short\text{-}tail} = \beta_{Long\text{-}tail}$	0.000	0.040	0.350	0.130		0.050
Beta (Market Value Weighted)						
Short-tail Lines	1.656	1.400	1.110	1.075	1.310	1.352
	(0.149)	(0.119)	(0.161)	(0.212)		(0.078)
Long-tail Lines	0.525	0.587	0.589	0.925	0.657	0.609
	(0.075)	(0.068)	(0.094)	(0.119)		(0.043)
F-Test: $\beta_{Short\text{-}tail} = \beta_{Long\text{-}tail}$	35.69 ***	28.11 ***	6.08 **	0.290		54.24 ***
Cost of Equity Capital (Equally Weighted)						
Short-tail Lines	13.9%	14.0%	13.4%	11.8%	13.3%	13.3%
Long-tail Lines	13.7%	13.5%	12.0%	12.7%	13.0%	13.0%
Cost of Equity Capital (Market Value Weighted)						
Short-tail Lines	19.9%	17.8%	15.3%	15.0%	17.0%	17.4%
Long-tail Lines	10.3%	10.9%	10.9%	13.7%	11.5%	11.1%

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 8

Full Information Fama-French 3-Factor Estimates with Sum Beta Adjustment: Short-Tail vs. Long-Tail Lines

Table displays full information beta estimates for the Fama-French 3 Factor Model by-line of property-liability insurance. Short-tail lines of insurance include all property coverages (fire, allied, etc.), financial and mortgage guaranty, fidelity and surety, etc. Long-tail lines of insurance include all liability coverages (products liability, other liability, etc.) as well as reinsurance, international, etc. See footnote 17 for a complete categorization of lines. The full-informaton beta comes from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is either the excess market coefficient, the SMB coefficient, or the HML coefficient estimated using equation (8) for firm i, bfullj is the estimated full-information coefficient estimate for industry (line of business) j, w_{ij} is the percent of firm i's net sales in industry (line of business) j. The regression is estimated via seemingly unrelated regressions both equally weighted and via weighted least squares (market weighted). The latter is used to obtain market-capitalization weighted industry full-informationbetas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Firms with estimated factor coefficients greater than 5 or less than -5 is removed from the sample. The full-information regression is estimated separately for each calendar year and as a pooled regression across all four years. The risk-free rate of interest is the 30-day T-bill rate in December 2000, 5.88 percent. The long-run historical premia for the excess market return, the SMB factor, and the HML factor as of December 2000 were 8.49 percent, 2.21 percent, and 4.63 percent, respectively.

	1997	1998	1999	2000	Average	Panel Estimate
Market Value Weighted Estimates						
Market Systematic Risk Factor						
Short-tail Lines	1.856	1.694	1.467	1.612	1.657	1.680
	(0.149)	(0.128)	(0.164)	(0.212)		(0.079)
Long-tail Lines	0.525	0.713	0.926	1.183	0.837	0.751
	(0.095)	(0.095)	(0.120)	(0.153)		(0.056)
SMB Factor						
Short-tail Lines	-0.233	0.029	0.300	0.026	0.030	0.058
	(0.222)	(0.193)	(0.186)	(0.250)		(0.106)
Long-tail Lines	-0.515	-0.423	-0.252	-0.218	-0.352	-0.426
	(0.142)	(0.143)	(0.136)	(0.180)		(0.074)
HML Factor						
Short-tail Lines	0.757	1.054	1.338	1.514	1.166	1.159
	(0.236)	(0.256)	(0.246)	(0.360)		(0.138)
Long-tail Lines	-0.086	0.498	0.814	0.890	0.529	0.380
	(0.151)	(0.190)	(0.180)	(0.259)		(0.097)
F-Test: $\beta_{Short\text{-}tail} = \beta_{Long\text{-}tail}$	48.01 ***	25.51 ***	4.79 **	0.40		61.49 ***
F-Test: $s_{Short\text{-}tail} = s_{Long\text{-}tail}$	0.76	2.39	3.90 **	0.40		9.35 ***
F-Test: $h_{Short\text{-}tail} = h_{Long\text{-}tail}$	6.04 ***	2.05	2.00	1.28		14.25 ***
Cost of Equity Capital (Equally Weighted)						
Short-tail Lines	16.3%	20.4%	21.0%	20.2%	19.5%	19.4%
Long-tail Lines	17.4%	17.2%	15.1%	19.4%	17.3%	17.3%
F-Test: $r_{Short\text{-}tail} = r_{Long\text{-}tail}$	0.06	0.71	3.55 *	0.05		1.66
Cost of Equity Capital (Market Value Weighted)						
Short-tail Lines	24.6%	25.2%	25.2%	26.6%	25.4%	25.6%
Long-tail Lines	8.8%	13.3%	17.0%	19.6%	14.7%	13.1%
F-Test: $r_{Short\text{-}tail} = r_{Long\text{-}tail}$	27.58 ***	12.77 ***	6.75 ***	2.69 *		53.79 ***

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 9

Full Information CAPM Beta Estimates with Sum Beta Adjustment: Personal Lines vs. Commericial Lines

Table displays full information CAPM beta estimates for personal lines and commericial lines property-liability insurance controlling for non-synchronous trading. Personal lines includes all net premiums written in homeowners, farmowners, earthquake, personal automobile iability, liability, and automobile physical damage. All others lines were considered commerical lines. The full-informaton beta comes from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is the equity beta estimated using equation (6) for firm i, $\beta full_j$ is the estimated full-information beta for industry (line of business) j, w_{ij} is the percent of firm i's net sales in industry (line of business) j. The regression is estimated using OLS and via weighted least squares to obtain market-capitalization weighted industry (line of business) full-information betas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated beta greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated separately for each calendar year and as a pool regression across all four years. The risk-free rate of interest is the 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical market risk premium as of December 2000 was 8.49 percent.

	1997	1998	1999	2000	Average	Panel Estimate
Beta (Equally Weighted)						
Personal Lines	1.114	1.007	0.779	0.734	0.908	0.929
	(0.263)	(0.276)	(0.256)	(0.258)		(0.132)
Commercial Lines	0.839	0.853	0.778	0.778	0.812	0.813
	(0.156)	(0.154)	(0.140)	(0.149)		(0.075)
F-Test: $\beta_{Personal} = \beta_{Commercial}$	0.630	0.180	0.000	0.020		0.450
Beta (Market Value Weighted)						
Personal Lines	0.749	0.736	0.446	0.725	0.664	0.686
	(0.103)	(0.091)	(0.137)	(0.187)		(0.061)
Commercial Lines	0.876	0.875	0.955	1.132	0.959	0.914
	(0.085)	(0.079)	(0.106)	(0.139)		(0.049)
F-Test: $\beta_{Personal} = \beta_{Commercial}$	0.690	1.040	6.50 ***	2.200		6.58 **
Cost of Equity Capital (Equally Weighted)						
Personal Lines	15.3%	14.4%	12.5%	12.1%	13.6%	13.8%
Commercial Lines	13.0%	13.1%	12.5%	12.5%	12.8%	12.8%
Cost of Equity Capital (Market Value Weighted)						
Personal Lines	12.2%	12.1%	9.7%	12.0%	11.5%	11.7%
Commercial Lines	13.3%	13.3%	14.0%	15.5%	14.0%	13.6%

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 10

Full Information Fama-French 3-Factor Estimates with Sum Beta Adjustment: Personal Lines vs. Commercial Lines

Table displays full information coefficient estimates for the Fama-French 3-Factor model for personal lines and commericial lines property-liability insurance. Personal lines include all net premiums written in homeowners, farmowners, earthquake, personal auto liability, and automobile physical damage. All other lines of business wei considered commerical lines. The full-informaton betas come from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_j$$

where β_i is either the excess market coefficient, the SMB coefficient, or the HML coefficient estimated using equation (8) for firm i, bfullj is the estimated full-informatior coefficient estimate for industry (line of business) j, w_{ij} is the percent of firm i's net sales in industry (line of business) j. The regression is estimated via seemingly unrelated regressions both equally weighted and via weighted least squares (market weighted). The latter is used to obtain market-capitalization weighted industry full-information betas. The weight is equal to the market capitalization of firm i relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Firms with estimated factor coefficients greaterthan 5 or less than -5 is removed from the sample. The full-information regression is estimated separately for each calendar year and as a pooled regression across allfour years.The risk-free rate of interest is the 30-day T-bill rate in December 2000, 5.88 percent. The long-run historical premia for the excess market return, the SMB factor, and the HML factor as of December 2000 were 8.49 percent, 2.21 percent, and 4.63 percent, respectively.

	1997	1998	1999	2000	Average	Panel Estimate
Market Value Weighted						
Market Systematic Risk Factor						
Personal Lines	0.886	1.029	0.980	1.263	1.040	0.999
	(0.102)	(0.098)	(0.140)	(0.180)		(0.061)
Commercial Lines	1.080	1.141	1.252	1.416	1.222	1.178
	(0.084)	(0.084)	(0.108)	(0.134)		(0.049)
Small - Big Capitalization Factor						
Personal Lines	-0.575	-0.467	-0.114	-0.189	-0.336	-0.430
	(0.153)	(0.147)	(0.159)	(0.213)		(0.081)
Commercial Lines	-0.305	-0.095	0.006	-0.104	-0.124	-0.120
	(0.126)	(0.127)	(0.122)	(0.158)		(0.066)
High - Low Book/Market Factor						
Personal Lines	0.042	0.619	1.330	1.465	0.864	0.632
	(0.162)	(0.196)	(0.210)	(0.307)		(0.106)
Commercial Lines	0.338	0.764	0.805	0.906	0.704	0.697
	(0.134)	(0.168)	(0.162)	(0.228)		(0.086)
F-Test: $\beta_{Personal} = \beta_{Commercial}$	1.66	0.60	1.80	0.33		4.02 **
F-Test: $s_{Personal} = s_{Commercial}$	1.44	2.88 *	0.27	0.07		6.72 ***
F-Test: $h_{Personal} = h_{Commercial}$	1.53	0.25	0.16	1.53		0.17
Cost of Equity Capital (Equally Weighted)						
Personal Lines	22.5%	22.3%	22.5%	25.3%	22.1%	23.1%
Commercial Lines	18.2%	20.3%	18.6%	21.2%	19.0%	19.5%
F-Test: $r_{Personal} = r_{Commercial}$	1.05	0.90	0.79	0.87		2.57
Cost of Equity Capital (Market Value Weighted)						
Personal Lines	12.3%	16.5%	20.1%	23.0%	18.0%	16.3%
Commercial Lines	15.9%	18.9%	20.3%	21.9%	19.2%	18.8%
F-Test: $r_{Personal} = r_{Commercial}$	2.95 *	0.96	0.00	0.10		3.75 *

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 11

Full Information CAPM Beta Estimates: Auto vs. Worker's Compenstion vs. All Other Property-Liability Lines of Insurance

Table displays full information CAPM beta estimates by-line of property-liability insurance. Automobile insurance includes personal automobile liability, commercial automobile liability, and automobile physical damage insurance.The full-informaton beta comes from the following cross-sectional regression

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_i$$

where β_i is the equity beta estimated using equation (6) for firm i, $\beta full_j$ is the estimated full-information beta for industry (line of business) j, W_{ij} is the percent of firm i's net sales in industry (line of business) j. The regression is estimated using OLS and via weighted least squares so we can obtain market-capitalization weighted industry (line of business) full-information betas. The weight is equal to the market capitalization of firm I relative to the market capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated beta greater than 5 or less than -5 is removed from the sample. The full-information regression was estimated separately for each calendar year and as a pool regression across all four years. The risk-free rate of interest used to estimate the cost of equity capital was the 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical market risk premium as of December 2000 was 8.49 percent.

	1997	1998	1999	2000	Average	Panel Estimate
Beta (Equally Weighted)						
Automobile Insurance	1.119	1.180	0.568	0.817	0.921	0.906
	(0.250)	(0.263)	(0.249)	(0.250)		(0.127)
Worker's Compensation	0.543	0.826	1.095	0.986	0.862	0.872
	(0.596)	(0.585)	(0.444)	(0.499)		(0.265)
All Other P&L Lines of Insurance	0.863	0.840	0.815	0.713	0.808	0.808
	(0.180)	(0.175)	(0.162)	(0.171)		(0.086)
F-Test: $\beta_{Auto} = \beta_{Worker's\ comp} = \beta_{All\ others}$	0.470	0.130	0.530	0.140		0.17
Beta (Market Value Weighted)						
Automobile Insurance	0.655	0.691	0.441	0.731	0.630	0.636
	(0.114)	(0.098)	(0.153)	(0.204)		(0.066)
Worker's Compensation	0.716	0.865	0.860	0.953	0.849	0.882
	(0.497)	(0.454)	(0.546)	(0.883)		(0.282)
All Other P&L Lines of Insurance	0.962	0.909	0.954	1.139	0.991	0.9486
	(0.104)	(0.091)	(0.127)	(0.174)		(0.059)
F-Test: $\beta_{Auto} = \beta_{Worker's\ comp} = \beta_{All\ others}$	1.520	1.090	2.54 *	0.850		4.97 ***
Cost of Equity Capital (Equally Weighted)						
Automobile Insurance	15.4%	15.9%	10.7%	12.8%	13.7%	13.6%
Worker's Compensation	10.5%	12.9%	15.2%	14.2%	13.2%	13.3%
All Other P&L Lines of Insurance	13.2%	13.0%	12.8%	11.9%	12.7%	12.7%
F-Test: $r_{Auto} = r_{Worker's\ Comp}$	0.710	0.080	0.960	0.080		0.010
F-Test: $r_{Auto} = r_{All\ Others}$	0.550	0.250	0.540	0.090		0.320
F-Test: $r_{Worker's\ Comp} = r_{Worker's\ Comp}$	0.240	0.000	0.330	0.250		0.050
Cost of Equity Capital (Market Value Weighted)						
Automobile Insurance	11.4%	11.7%	9.6%	12.1%	11.2%	11.3%
Worker's Compensation	12.0%	13.2%	13.2%	14.0%	13.1%	13.4%
All Other P&L Lines of Insurance	14.0%	13.6%	14.0%	15.5%	14.3%	13.9%
F-Test: $r_{Auto} = r_{Worker's\ Comp}$	0.010	0.140	0.530	0.060		0.700
F-Test: $r_{Auto} = r_{All\ Others}$	3.03 *	2.080	4.87 **	1.620		9.45 ***
F-Test: $r_{Worker's\ Comp} = r_{Worker's\ Comp}$	0.200	0.010	0.020	0.040		0.050

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.

Table 12

Full Information Fama-French 3-Factor Estimates with Sum Beta Adjustment: Auto vs. Workers' Compensation vs. All Other P&L Lines

Table displays full information beta estimates for the Fama-French 3 Factor Model by-line of property-liability insurance. Automobile insurance includes personal automobile liability commercial automobile liability, and automobile physical damage. The full-information beta comes from the following cross-sectional regression:

$$\beta_i = \Sigma\, \beta full_j(W_{ij}) + \upsilon_i$$

where β_i is either the excess market coefficient, the SMB coefficient, or the HML coefficient estimated using equation (8) for firm i, $\beta full_j$ is the estimated full-information coefficient estimate for industry (line of business) j, W_{ij} is the percent of firm i's net sales in industry (line of business) j. The regression is estimated via seemingly unrelated regressions both equally weighted and via weighted least squares (market weighted). The latter is used so we can obtain market-capitalization weighted industry full-information betas. The weight is equal to the market capitalization of firm i relative to the total capitalization of all NYSE, AMEX, and Nasdaq stocks. Any firm with an estimated factor coefficient greater than 5 or less than -5 is deleted from the sample. The full-information regression was estimated separately for each calendar year and as a pooled regression across all four years. The risk-free rate of ineterest used to estimate the cost of equity capital was the 30 day T-bill rate in December 2000, 5.88 percent. The long-run historical premium for the excess market return, the SMB factor, and the HML factor as of December 2000 were 8.49 percent, 2.21 percent, and 4.63 percent, respectively.

	1997	1998	1999	2000	Average	Panel Estimate
Market Value Weighted Results						
Beta Factor						
Automobile Insurance	0.795	1.016	1.023	1.284	1.029	0.971
	(0.112)	(0.105)	(0.156)	(0.197)		(0.066)
Worker's Compensation	0.906	0.917	0.815	0.576	0.804	0.867
	(0.493)	(0.485)	(0.557)	(0.853)		(0.283)
All Other P&L Lines of Insurance	1.164	1.171	1.263	1.484	1.271	1.228
	(0.103)	(0.097)	(0.129)	(0.168)		(0.059)
Small - Big Capitalization Factor						
Automobile Insurance	-0.470	-0.465	0.129	-0.258	-0.266	-0.438
	(0.168)	(0.158)	(0.176)	(0.233)		(0.089)
Worker's Compensation	-0.610	0.552	0.479	0.495	0.229	0.070
	(0.735)	(0.731)	(0.630)	(1.009)		(0.379)
All Other P&L Lines of Insurance	-0.360	-0.170	0.053	-0.116	-0.148	-0.143
	(0.153)	(0.146)	(0.146)	(0.199)		(0.079)
High - Low Book/Market Factor						
Automobile Insurance	0.135	0.727	0.146	1.447	0.614	0.687
	(0.179)	(0.210)	(0.233)	(0.335)		(0.115)
Worker's Compensation	0.154	0.758	0.146	-0.055	0.251	0.155
	(0.782)	(0.971)	(0.834)	(1.450)		(0.494)
All Other P&L Lines of Insurance	0.277	0.673	0.898	1.030	0.720	0.712
	(0.163)	(0.194)	(0.194)	(0.286)		(0.103)
F-Test: $\beta_{Auto} = \beta_{Worker's\ Comp} = \beta_{All\ Others}$	2.26	0.50	0.63	0.53		3.29 **
F-Test: $s_{Auto} = s_{Woorker's\ Comp} = s_{All\ Oothers}$	0.11	1.55	1.15	0.34		3.04 **
F-Test: $h_{Auto} = h_{Worker's\ Comp} = h_{All\ Others}$	0.13	0.01	1.44	0.83		0.55
Cost of Equity Capital (Equally Weighted)						
Automobile Insurance	22.9%	23.3%	19.0%	24.6%	22.4%	22.3%
Worker's Compensation	20.5%	20.0%	18.2%	20.1%	19.7%	19.6%
All Other P&L Lines of Insurance	17.3%	20.1%	20.4%	21.8%	19.9%	19.8%
F-Test: $r_{Auto} = r_{Worker's\ Comp}$	0.08	0.06	0.01	0.33		0.44
F-Test: $r_{Auto} = r_{All\ Other\ P\text{-}L\ Lines}$	1.68	0.20	0.09	0.38		1.24
F-Test: $r_{AWorkers'\ Comp} = r_{All\ Other\ P\text{-}L\ Lines}$	0.16	0.00	0.11	0.05		0.00
Cost of Equity Capital (Market Value Weighted)						
Automobile Insurance	12.2%	16.8%	15.5%	22.9%	16.9%	16.3%
Worker's Compensation	12.9%	18.4%	14.5%	11.6%	14.4%	14.1%
All Other P&L Lines of Insurance	16.2%	18.6%	20.9%	23.0%	19.7%	19.3%
F-Test: $r_{Auto} = r_{Worker's\ Comp}$	0.01	0.03	0.55	0.93		0.27
F-Test: $r_{Auto} = r_{All\ Other\ P\text{-}L\ Lines}$	2.75 *	0.39	0.06	0.00		4.01 **
F-Test: $r_{AWorkers'\ Comp} = r_{All\ Other\ P\text{-}L\ Lines}$	0.19	0.00	0.66	0.84		1.35

- ***, **, * significant at the 1, 5, or 10 percent level, respectively. Standard errors in parantheses.